CRYSTALS OF AN AURORA-A TPX2 COMPLEX, TPX2 BINDING SITE OF AURORA-A, AURORA-A LIGANDS AND THEIR USE

The present invention relates to crystals of phosphorylated Aurora-A kinase fragment alone and in complex with a ligand, amino acid residues 1-43 of human TPX2. This invention also relates to methods for designing and selecting ligands, in particular allosteric inhibitors of Aurora-A, that bind to the Aurora-A kinase and their use. Further, the present invention relates to certain indene and indole derivatives.

At the beginning of mitosis, eukaryotic cells undergo a dramatic reorganization. The nuclear envelope breaks down and microtubules rearrange around chromatin into a bipolar spindle structure that carries out the duty of separating sister chromatids. Chromatin signals to the spindle assembly machinery through Ran, a small Ras-like GTPase that is concentrated in its GTP-bound form around chromatin. Ran function in spindle assembly is connected to its role in nucleocytoplasmic transport. RanGTP releases crucial spindle assembly factors such as TPX2 and NuMA from the transport factors that mediate their import into the nucleus at interphase. After nuclear envelope breakdown, the presence of free TPX2 in the vicinity of chromatin is thought to nucleate microtubules that are subsequently organised into a spindle by microtubule motors. Additionally, TPX2 localises an essential mitotic kinase, Aurora-A, to spindle microtubules (Kufer, T. A., Sillje, H. H., Korner, R., Gruss, O. J., Meraldi, P., and Nigg, E. A. (2002). Human TPX2 is required for targeting Aurora-A kinase to the spindle. J. Cell Biol. *158*, 617-623).

Aurora kinases constitute a family of serine-threonine protein kinases whose localization and activities are precisely choreographed as a cell progresses through mitosis. Aurora-A is a cell-cycle regulated serine-threonine kinase involved in chromosome segregation and cytokinesis (Bischoff, J.R. and Plowman, G.D. (1999). The Aurora/Ipl1p kinase family: regulators of chromosome segregation and cytokinesis. Trends Cell Biol. 9, 454-459). It plays a major role in cell-cycle progression and has also been described as an oncogene. It maps to a chromosome region frequently amplified in tumours (Dutertre, S. et al. (2002). On the role of Aurora A in centrosome function. Oncogene 21, 6175-6183). It is overexpressed in a variety of human tumours, in particular breast and colon cancer, but has limited expression in normal tissues (Sen, S. et al. (1997). A putative serine/threonine kinase encoding gene BTAK on chromosome 20q13 is amplified and overexpressed in human breast cancer cell lines. Oncogene 14, 2195-2200; Bischoff, J.R. et al. (1998). A homologue of Drosophila Aurora kinase is oncogenic and amplified in human colorectal cancers. EMBO J. 17, 3052-3065). Overexpression of active Aurora-A transforms rat fibroblasts so that they

are capable of growing tumours in nude mice, while an inactive mutant is unable to cause oncogenic transformations. As an oncogenic protein kinase, Aurora-A is a target for the development of specific inhibitors that may be useful as cancer therapeutics. Despite the importance of Aurora-A for both cell division and cancer perspectives, little is known at present about its downstream targets and activation/deactivation mechanisms.

Several factors contribute to the activity of a serine/threonine kinase. These include the proper positioning of active site residues and the correct organisation of the substrate-binding site (the "activation segment"). Phosphorylation of a threonine residue within the activation segment is often required to elicit kinase activity and Aurora-A is no exception. Phosphorylation of a threonine in the Aurora-A activation segment ($Thr288^{AUR}$, human numbering) is crucial for activity, although it is unclear as to whether it is catalysed *in vivo* by an upstream kinase or by Aurora-A itself (Bischoff, et al. (1999) *supra*). Structural studies of c-AMP dependent protein kinase (cAPK) have shown that when the corresponding threonine residue ($Thr197^{cAPK}$) is phosphorylated, the activation segment is in an active conformation. Cyclin-dependent kinases (CDKs) require not only phosphorylation of the equivalent threonine ($Thr160^{CDK}$) but also the binding by a partner protein, cyclin-A, to be fully activated. Aurora-A might also rely on a similar mechanism. It has recently been reported that in vertebrates the interaction of Aurora-A with a partner protein, TPX2, leads to a strong activation of the kinase. Upon TPX2 binding, the *in vitro* autophosphorylation activity of Aurora-A is increased and dephosphorylation is prevented (Kufer et al., (2002) *supra*).

ATP competitive inhibitors that are specific for different kinases are used as therapeutic agents in cancer treatment (Garcia-Echeverria, C., et al. (2000). ATP site-directed competitive and irreversible inhibitors of protein kinases. Med. Res. Rev. 20, 28-57). Although ATP-binding sites at the moment are the most common targets for the design of kinase inhibitors, it is difficult to achieve selectivity of such inhibitors due to the similarity in kinase active sites, which only have minor differences of surrounding amino-acid residues (Cheetham, G.M.T. et al. (2002). Crystal structure of Aurora-2, an oncogenic serine-threonine kinase. J. Biol. Chem. 277, 42419-42422).

The structure of unphosphorylated Aurora-A has been previously reported (Cheetham, G.M.T. et al. (2002). Crystal structure of Aurora-2, an oncogenic serine-threonine kinase. J. Biol. Chem. 277, 42419-42422; Nowakowski, J. et al. (2002). Structures of the cancer-related Aurora-A, FAK, and EphA2 protein kinases from nanovolume crystallography. Structure 10, 1659-1667). Further, WO 03/031606 describes the 3-dimensional crystal structure of the kinase catalytic domain of Aurora-A in a complex with the ATP analogue ATP-PNP, and the

3-dimensional crystal structure of the kinase catalytic domain of Aurora-A in complex with a synthetic inhibitor. Aurora-A has the typical three-dimensional structure of protein kinases, with the active site situated between the N- and C-terminal lobe. Binding of ATP involves amino-acid residues that are conserved among all kinases. Extensive structural work has shown that kinases in their active state all assume a similar structural framework, with the 'activation segment' in a similar conformation competent for substrate binding (Huse, M. and Kuriyan, J. (2002). The conformational plasticity of protein kinases. Cell 109, 275-282). However, they differ in the molecular mechanisms to achieve such an active form. In the case of Abl kinase, subtle differences in its activation mechanisms have been exploited with the development of the Abl-specific inhibitor Gleevec, which is used as a leukaemia therapeutic agent (Capdeville, R. et al. (2002). Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug. Nat. Rev. Drug Discov. 1, 493-502).

In the case of Aurora-A, activation is achieved by both phosphorylation and by the binding of a specific activator, the protein TPX2 (Eyers, P.A., et al. (2003). A novel mechanism for activation of the protein kinase Aurora-A. Curr. Biol. 13, 691-697). Blocking this activator-binding site would provide a means to downregulate this kinase specifically.

In view thereof, it was an object of the present invention to elucidate the structure of the Aurora-A TPX2 binding site, to provide means for identifying compounds that bind to Aurora-A and preferably modulate Aurora-A activity, and to provide such compounds.

The present invention relates to (a) crystals of a fragment of phosphorylated human Aurora-A kinase alone (amino acid residues 122-403; hereinafter referred to as Aurora-A(ΔN)), and (b) crystals of said fragment of phosphorylated human Aurora-A kinase in complex with a ligand, i. e. an Aurora-A ligand complex. The Aurora-A ligand is a fragment of TPX2 which is a minimal activating domain of TPX2. This minimal activating domain of TPX2 consists of amino acid residues 1-43 of human TPX2 and hereinafter is referred to as TPX2(1-43).

The present invention provides the structure coordinates of the phosphorylated human Aurora-A(ΔN) kinase. The complete coordinates are listed in Table A.

The present invention also provides the structure coordinates of the phosphorylated human Aurora-A(ΔN)/TPX2(1-43) complex. The complete coordinates are listed in Table B.

The present invention also describes a method for determining at least a portion of the three-dimensional structure of molecules or molecular complexes which contain at least some

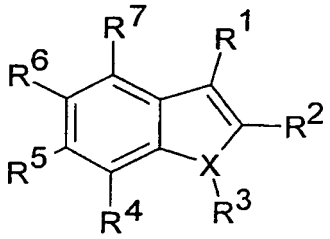
structurally similar features to the Aurora-A TPX2 binding domain. It is preferred that these molecules or molecular complexes comprise at least a part of the ligand binding site defined by structure coordinates of Aurora-A amino acids Q127, W128, R126, L159, F157, E170, L169, V206, Y199, H187, R179, L178, V182, Y199, L188, I184, V252, K250, P282, H280 according to Table B, or a mutant or homologue thereof. The numbering system as used herein refers to the protein sequences for human Aurora-A.

The present invention also provides a machine-readable data storage medium which comprises a data storage material encoded with machine readable data defined by the structure coordinates of phosphorylated human Aurora-A(ΔN) kinase according to Table A or a homologue thereof, or of the phosphorylated human Aurora-A(ΔN)/TPX2(1-43) complex according to Table B.

The present invention further provides a binding site in Aurora-A for an Aurora-A ligand such as TPX2 or fragments thereof, as well as methods for designing or selecting further Aurora-A ligands and in particular Aurora-A modulators including agonists, partial agonists, antagonists, partial antagonists of Aurora-A using information about the crystal structures disclosed herein.

The present invention further provides allosteric inhibitors of Aurora-A, wherein at least a portion of the inhibitor binds with any portion or all of residues Q127, W128, R126, L159, F157, E170, L169, V206, Y199, H187, R179, L178, V182, Y199, L188, I184, V252, K250, P282, H280 of Aurora-A according to Table B.

The present invention in particular relates to indole and indene derivatives of formula (I)



(I)

wherein

R^1 represents hydrogen, alkylene-COR^{11}, alkylene-NHR^8, alkylene-OR^8, or alkylene-SR^8;

R^2 represents hydrogen, alkylene-COR^{11}, alkylene-NHR^8, alkylene-OR^8, or alkylene-SR^8;

R^3 represents hydrogen, alkyl, alkylene-R^9, alkenylene-R^9, alkynylene-R^9, or arylene-R^9;

R^4 represents hydrogen;

R^5 represents hydrogen, alkyl, OR^{10}, NHR^{10}, SR^{10}, alkylene-R^{10}, alkenylene-R^{10}, alkynylene-R^{10}, or arylene-R^{10};

R^6 represents hydrogen, alkyl, OR^{10}, NHR^{10}, SR^{10}, alkylene-R^{10}, alkenylene-R^{10}, alkynylene-R^{10}, or arylene-R^{10};

R^7 represents hydrogen;

R^8 represents hydrogen, CO-alkyl, $(aa)_m asp(aa)_n$, $(aa)_m glu(aa)_n$, or $(aa)_m cys(aa)_n$, or optionally substituted alkyl, aryl or heteroaryl;

R^9 represents NH-alkyl, $N(alkyl)_2$, $N^+(alkyl)_3$, optionally substituted aryl, or optionally substituted heteroaryl;

R^{10} represents hydrogen or a mono- or bicyclic, saturated, partially unsaturated or aromatic, alicyclic or heterocyclic radical which may be substituted;

R^{11} represents hydrogen, alkyl or haloalkyl;

X represents a nitrogen atom or CH;

aa represents an amino acid radical; and

n is zero or an integer of 1 to 10;

m is zero or an integer of 1 to 10,

provided that R^1 and R^2 are not both hydrogen and that R^5 and R^6 are not both hydrogen, and optical isomers, physiologically acceptable salts, derivatives and prodrugs thereof.

The present invention also relates to the pharmaceutical compostions containing Aurora-A ligands, such as said indole and indene derivatives, and the use of Aurora-A ligands, such as said indole and indene derivatives, in therapy, in particular in cancer treatment.

In the drawings the figures show as

(figure 1) (A) anti-*Xenopus* TPX2 Western blot (upper panel) and anti-*Xenopus* Aurora-A Western blot (lower panel) of GST (glutathione S transferase) (lanes 1 and 2), GST *Xenopus* TPX2 (lanes 3 and 4) or GST *Xenopus* TPX2(1-39) (lanes 5 and 6) which proteins were incubated in *Xenopus* cytostatic factor (CSF) arrested extracts in the presence or absence of RanQ69L-GTP and immunoprecipitated with GST antibody-coated beads;
(B) anti-GST Western blot (upper panel) and anti-human Aurora-A Western blot (lower panel) of GST (lane 1), GST TPX2(1-43) (lane 2) or GST TPX2(15-43) (lane 3) proteins that were incubated in mitotic HeLa cell extract and immunoprecipitated with GST antibody-coated beads;
(C) autoradiography of the SDS-PAGE gel (left panel) and the corresponding Coomassie-stained gel (right panel) after *in vitro* phosphorylation (γ^{32}P-ATP)

of histone H3 by human Aurora-A in the presence of full-length TPX2, GST TPX2(1-43) or GST TPX2(15-43) (lanes 2, 3 and 4 respectively);
(D) much of the phosphorylation signal in GST after *in vitro* phosphorylation (γ^{32}P-ATP) of TPX2(1-43) by Aurora-A (lane 1) followed by TEV cleavage (lane 2);
(E) an anti-human Aurora-A Western blot (upper panel) and an anti-phosphoAurora-A Western blot (lower panel) after phosphatase PP1 treatment of human Aurora-A in the absence or presence of full-length TPX2, GST TPX2(1-43) or GST TPX2(15-43) followed by detection of Aurora-A by a polyclonal antibody (upper panel) and an antibody specific for Aurora-A phosphorylated at Thr288^{AUR} (lower panel);

(figure 2) (A) an *in vitro* pull-down assay with respect to the binding of full-length Aurora-A, AuroraΔN or AuroraΔN(D274N) to GST TPX2(1-43) (lanes 4, 5 and 6), and to GST (lanes 1, 2 and 3);
(B) an anti-phospho Aurora-A Western blot of wild-type Aurora(ΔN) and the D274N mutant when expressed in *E. coli.*
(C) autoradiograph for detecting i*n vitro* phosphorylation of histone H3 by Aurora(ΔN) in the presence or absense of TPX2(1-43) (lane 2 compared to lane 1), the cleaved (lane 2) or uncleaved (lane 3) GST TPX2(1-43) fusion protein;
(D) Sequence alignment of TPX2 N-terminal domain from human (*H*), *Xenopus* (*X*) and puffer fish (*F*), secondary structure elements being shown above the sequences *in red* (upstream extended stretch) and *pink* (downstream helical stretch), and intervening residues not modelled being marked with a dotted line;
(E) Sequence alignment of Aurora-A kinase catalytic domain from three vertebrate species that contain TPX2 (human, *H.AUR-A*; *Xenopus*, *X.AUR-A*; puffer fish, *F.AUR-A*), two invertebrates that do not contain TPX2 (*Drosophila, D.AUR-A; C.elegans*, *C.AUR-A*) together with human and *Xenopus* Aurora-B (*H.AUR-B*, *X.AUR-B*) and vertebrate cAPK, wherein Aurora-A secondary structural elements are labelled above the alignment, the phosphorylated Thr288 (human numbering) is shown, and residues that interact with 7-21^{TPX} or 30-43^{TPX} being indicated by filled or open circles respectively;

(figure 3) ribbon style drawings of the structure of Aurora-A bound to TPX2 as
(A) a view of the complex between the catalytic domain of human Aurora (AuroraΔN) and the N-terminal domain of TPX2 shown in typical kinase orientation, an upstream stretch of TPX2 binding at the N-terminal lobe of

Aurora-A, and a downstream stretch binding between the two lobes, and a dotted line marking the approximate path of the linker connecting the two TPX2 stretches (disordered and not modeled);

(B) a view of the complex after a 180° rotation about the vertical axis in respect to view in panel A showing more clearly the two stretches of TPX2 that bind to Aurora-A;

(C) the upstream stretch of TPX2 (residues 7-21^{TPX}) that binds at a hydrophobic surface groove present in the N-terminal lobe of the kinase, wherein details of the extensive interactions are shown in the same orientation as in panel B;

(D) the downstream helical stretch of TPX2 (residues 30-43^{TPX}) that binds Aurora-A near helix αC and the activation segment, close to but not directly in contact with phospho-Thr288^{AUR}, wherein details of interactions being shown in the same orientation as in panel B and C.

(figure 4) ribbon style drawings of conformational states of phosphorylated Aurora-A in the presence and absence of TPX2 as

(A) an overlay showing that the structures of Aurora-A when bound to TPX2 and when unbound are closely superposable at the position of active site residues and of helix αC, but diverge at the activation segment between residues His280^{AUR} and Leu293^{AUR}, wherein Phospho-Thr288^{AUR} points inwards in the TPX2-bound structure and outwards in the kinase alone structure;

(B) an illustration of conformational changes upon TPX2 binding, according to which the activation segments of the kinase alone structure (left panel) and of the TPX2-bound structure (right panel) are shown in a view rotated by approximately 90° with respect to panel A, and TPX2 binding results in the reorganization of the activation segment, with a 10 Å movement of Thr288^{AUR};

(C) a schematic representation of the molecular mechanism of TPX2-mediated activation of Aurora-A, according to which the upstream stretch of TPX2 anchors the regulator to the N-terminal lobe of the kinase and the downstream stretch (helix) hooks the activation segment triggering a lever-arm like movement, where rotations at His280^{AUR} and Pro282^{AUR} pull on Thr288^{AUR};

(figure 5) the structure coordinates of phosphorylated human Aurora-A(ΔN) kinase (table A);

(figure 6) the structure coordinates of phosphorylated human Aurora-A(ΔN)/TPX2(1-43) complex (table B).

Using X-ray crystallography, crystal structures of phosphorylated human Aurora-A(ΔN) alone at 2.75 Å resolution and in complex with a minimal activating domain of TPX2 at 2.5 Å resolution have been determined and the specific site of TPX2-mediated activation of Aurora-A has been found. TPX2 binds at two sites on the kinase. One stretch of TPX2 (residues 8-19) binds at the N-terminal lobe of the kinase and another stretch (31-38) binds between the two lobes at the phosphorylated activation segment, positioning it for substrate binding. The TPX2-binding site located on the N-terminal lobe of the kinase consists of an extended surface groove formed by Aurora-specific residues. The groove consists of two pockets, one shaped by residues E170, L169, V206, R179, L178, V182, Y199 and the other shaped by residues Q127, W128, R126, L159 and F157. The two pockets are occupied by hydrophobic side chains of TPX2, Y8, Y10, A12 and P13 on one side, and F16, I17 and F19 on the other. From the structure and biochemical data it is concluded that binding of this TPX2 stretch is necessary but not sufficient for activation of Aurora-A, allowing the downstream TPX2 stretch to bind with enough affinity to achieve activation. Binding of a small-molecule inhibitor to the TPX2-recognition groove on the N-terminal lobe of Aurora-A would decrease the activity of this kinase by blocking its specific activation mechanism.

The Aurora-A kinase and the TPX2 protein described herein are intended to include any polypeptide which has the activity of the naturally occurring Aurora-A kinase and TPX2 protein, respectively. Aurora-A and TPX2 contemplated herein include all vertebrate and mammalian forms such as rat, mouse, pig, goat, horse, guinea pig, rabbit, monkey, orangutan and human. Such terms also include polypeptides that differ from naturally occurring forms of Aurora-A kinase and TPX2 protein by having amino acid deletions, substitutions, and additions, but which retain the activity of Aurora-A kinase and TPX2 protein, respectively. Particular Aurora-A sequences are shown in figure 2E and particular TPX2 sequences are shown in figure 2D. According to the present invention, the human sequences are preferred.

The crystal structures of the invention preferably contains at least 25%, more preferably at least 50%, more preferably at least 75%, more preferably at least 90%, more preferably at least 95%, more preferably at least 99%, and most preferably all of the coordinates listed in Table A and Table B, respectively.

The crystal of the phosphorylated human Aurora-A(ΔN) kinase of the invention preferably has the following unit cell dimensions in angstroms: a = 81.18 ± 5%, b = 81.18 ± 5%, c = 169.62 ± 5% and the space group $P6_122$.

The crystal of the phosphorylated human Aurora-A(ΔN) kinase/TPX2(1-43) complex of the invention preferably has the following unit cell dimensions in angstroms: a = 59.63 ± 5%, b = 81.72 ± 5%, c = 83,05 ± 5% and an orthorhombic space group $P2_12_12_1$.

The three-dimensional structure of the phosphorylated human Aurora-A(ΔN) kinase and of the phosphorylated human Aurora-A(ΔN) kinase/TPX2(1-43) complex of this invention are defined by a set of structure coordinates as set forth in Table A and Table B, respectively. The term "structure coordinates" refers to Cartesian coordinates derived from mathematical equations related to the patterns obtained on diffraction of a monochromatic beam of X-rays by the atoms (scattering centers) of the protein or protein complex in crystal form. The diffraction data are used to calculate an electron density map of the repeating unit of the crystal. The electron density maps are then used to establish the positions of the individual atoms of the complex.

Those of skill in the art will understand that a set of structure coordinates for a kinase or a kinase/ligand complex or a fragment thereof, is a relative set of points that define a shape in three dimensions. Thus, it is possible that an entirely different set of coordinates could define a similar or identical shape. Moreover, slight variations in the individual coordinates will have little effect on overall shape.

The variations in coordinates discussed above may be generated because of mathematical manipulations of the structure coordinates. For example, the structure coordinates set forth in Table A could be manipulated by crystallographic permutations of the structure coordinates, fractionalization of the structure coordinates; integer additions or subtractions to sets of the structure coordinates, inversion of the structure coordinates or any combination of the above.

Alternatively, modifications in the crystal structure due to mutations, additions, substitutions, and/or deletions of amino acids, or other changes in any of the components that make up the crystal could also account for variations in structure coordinates. If such variations are within an acceptable standard error as compared to the original coordinates, the resulting three-dimensional shape is considered to be the same.

Various computational analyses are therefore necessary to determine whether a molecule or molecular complex or a portion thereof is sufficiently similar to all or parts of the kinase or the kinase/ligand complex described above as to be considered the same. Such analyses may be carried out in current software applications, such as the Molecular Similarity application of

QUANTA (Molecular Simulations Inc., San Diego, CA) version 4.1, and as described in the accompanying User's Guide.

The Molecular Similarity application permits comparisons between different structures, different conformations of the same structure, and different parts of the same structure. The procedure used in Molecular Similarity to compare structures is divided into four steps: 1) load the structures to be compared; 2) define the atom equivalences in these structures; 3) perform a fitting operation; and 4) analyze the results.

Each structure is identified by a name. One structure is identified as the target (i. e., the fixed structure); all remaining structures are working structures (i. e., moving structures). Since atom equivalency within QUANTA is defined by user input, for the purpose of this invention equivalent atoms are defined as protein backbone atoms (N, C and O) for all conserved residues between the two structures being compared. Also, only rigid fitting operations are considered.

When a rigid fitting method is used, the working structure is translated and rotated to obtain an optimum fit with the target structure. The fitting operation uses an algorithm that computes the optimum translation and rotation to be applied to the moving structure, such that the root mean square difference of the fit over the specified pairs of equivalent atom is an absolute minimum. This number, given in angstroms, is reported by QUANTA.

For the purpose of this invention, any molecule or molecular complex that has a root mean square deviation of conserved residue backbone atoms (N, C, O) of less than 1.5 A when superimposed on the relevant backbone atoms described by structure coordinates listed in Table A or Table B are considered identical. More preferably, the root mean square deviation is less than 1.0 A. In a preferred embodiment of the present invention, the molecule or molecular complex comprises at least a portion of the ligand binding site defined by structure coordinates of Aurora-A amino acids Q127, W128, R126, L159, F157, E170, L169, V206, Y199, H187, R179, L178, V182, Y199, L188, I184, V252, K250, P282, H280 according to Table B, or a mutant or homologue of said molecule or molecular complex. More preferred are molecules or molecular complexes comprising all or any part of the ligand binding site defined by structure coordinates of Aurora-A amino acids Y199, L178, W128, H187, L188, I184 according to Table B, or a mutant or homologue of said molecule or molecular complex.

The term "complex" or "molecular complex" means Aurora-A or a mutant or homologue of Aurora-A in a covalent or non-covalent association with a chemical entity or compound.

For purposes of the present invention, by "at least a portion of" it is meant all or any part of the ligand binding site defined by these structure coordinates.

By "mutant or homologue" as used herein it is meant a molecule or molecular complex having a similar structure and/or sequences to Aurora-A or TPX2. By "similar structure" it is meant a mutant or homologue having a binding pocket or binding domain that have a root mean square deviation from the backbone atoms of said Aurora-A or TPX2 amino acids of not more than 1.5 Angstroms. By "similar sequence" it is meant a mutant or homologue having 30%, or more preferably 75%, identity with Aurora-A or TPX2 over an amino acid sequnce of at least 30, preferably at least 50, in particular at least 100 and especially at least 200 consecutive amino acid.

The term "root mean square deviation" means the square root of the arithmetic mean of the squares of the deviations from the mean. It is a way to express the deviation or variation from a trend or object. For purposes of this invention, the "root mean square deviation" defines the variation in the backbone of a protein or protein complex from the relevant portion of the backbone of the Aurora-A fragment of the complex as defined by the structure coordinates described herein.

Once the structure coordinates of a protein crystal have been determined they are useful in solving the structures of other crystals.

Thus, in accordance with the present invention, the structure coordinates of the kinase of the kinase/ligand complex, and portions thereof is stored in a machine-readable storage medium. Such data may be used for a variety of purposes, such as drug discovery and x-ray crystallographic analysis of protein crystals.

Accordingly, in one embodiment of this invention is provided a machine-readable data storage medium comprising a data storage material encoded with the structure coordinates set forth in Table A and/or with the structure coordinates set forth in Table B.

For the first time, the present invention permits the use of structure-based or rational drug design techniques to design, select, and synthesize chemical entities, including inhibitory and stimulatory compounds that are capable of binding to Aurora-A, or any portion thereof.

One particularly useful drug design technique enabled by this invention is iterative drug design. Iterative drug design is a method for optimizing associations between a protein and a compound by determining and evaluating the three-dimensional structures of successive sets of protein/compound complexes.

Those of skill in the art will realize that association of natural ligands or substrates with the binding pockets of their corresponding kinases or enzymes is the basis of many biological mechanisms of action. The term "binding pocket" as used herein, refers to a region of a molecule or molecular complex, that, as a result of its shape, favorably associates with another chemical entity or compound. Similarly, many drugs exert their biological effects through association with the binding pockets of kinases and enzymes. Such associations may occur with all or any parts of the binding pockets. An understanding of such associations will help lead to the design of drugs having more favorable associations with their target kinase, and thus, improved biological effects. Therefore, this information is valuable in designing potential ligands or modulators, e.g. inhibitors, of Aurora-A kinase.

The term "associating with" refers to a condition of proximity between chemical entities or compounds, or portions thereof. The association may be non-covalent--wherein the juxtaposition is energetically favored by hydrogen bonding or van der Waals or electrostatic interactions--or it may be covalent.

In iterative drug design, crystals of a series of protein/compound complexes are obtained and then the three-dimensional structures of each complex is solved. Such an approach provides insight into the association between the proteins and compounds of each complex. This is accomplished by selecting compounds with inhibitory activity, obtaining crystals of this new protein/compound complex, solving the three-dimensional structure of the complex, and comparing the associations between the new protein/compound complex and previously solved protein/compound complexes. By observing how changes in the compound affected the protein/compound associations, these associations may be optimized.

In some cases, iterative drug design is carried out by forming successive protein-compound complexes and then crystallizing each new complex. Alternatively, a pre-formed protein crystal is soaked in the presence of a compound, thereby forming a protein/compound complex and obviating the need to crystallize each individual protein/compound complex.

As used herein, the term "soaked" refers to a process in which the crystal is transferred to a solution containing the compound of interest.

The structure coordinates set forth in Table A and Table B can also be used to aid in obtaining structural information about another crystallized molecule or molecular complex. This may be achieved by any of a number of well-known techniques, including molecular replacement.

The structure coordinates set forth in Table A and Table B can also be used for determining at least a portion of the three-dimensional structure of molecules or molecular complexes which contain at least some structurally similar features to Aurora-A. In particular, structural information about another crystallized molecule or molecular complex may be obtained. This may be achieved by any of a number of well-known techniques, including molecular replacement.

Therefore, in another embodiment this invention provides a method of utilizing molecular replacement to obtain structural information about a crystallized molecular complex whose structure is unknown comprising the steps of:

a) generating an X-ray diffraction pattern from said crystallized molecular complex;
b) applying at least a portion of the structure coordinates set forth in Table A and Table B to the X-ray diffraction pattern to generate a three-dimensional electron density map of the molecular complex whose structure is unknown; and
c) using all or a portion of the structure coordinates set forth in Table A and Table B to generate homology models of Aurora-A or any other kinase ligand binding domain.

Preferably, the crystallized molecular complex is obtained by soaking a crystal of this invention in a solution.

Molecular replacement provides an accurate estimation of the phases for an unknown structure. Phases are a factor in equations used to solve crystal structures that can not be determined directly. Obtaining accurate values for the phases, by methods other than molecular replacement, is a time-consuming process that involves iterative cycles of approximations and refinements and greatly hinders the solution of crystal structures. However, when the crystal structure of a protein containing at least a homologous portion has been solved, the phases from the known structure provide a satisfactory estimate of the phases for the unknown structure.

Thus, this method involves generating a preliminary model of a molecule or molecular complex whose structure coordinates are unknown, by orienting and positioning the relevant

portion of the Aurora-A ligand complex according to Table B within the unit cell of the crystal of the unknown molecule or molecular complex so as best to account for the observed X-ray diffraction pattern of the crystal of the molecule or molecular complex whose structure is unknown. Phases can then be calculated from this model and combined with the observed X-ray diffraction pattern amplitudes to generate an electron density map of the structure whose coordinates are unknown. This, in turn, can be subjected to any well-known model building and structure refinement techniques to provide a final, accurate structure of the unknown crystallized molecule or molecular complex [E. Lattman,"Use of the Rotation and Translation Functions", in Meth. Enzymol., 115, pp. 55-77 (1985); M. G. Rossmann, ed.,"The Molecular Replacement Method", Int. Sci. Rev. Set., No. 13, Gordon & Breach, New York (1972)].

The structure of any portion of any crystallized molecule or molecular complex, or mutant, or homologue that is sufficiently homologous to any portion of the Aurora-A ligand complex can be solved by this method.

The structure coordinates are also particularly useful to solve the structure of crystals of Aurora-A ligand co-complexed with a variety of chemical entities. This approach enables the determination of the optimal sites for interaction between chemical entities, including interaction of candidate Aurora-A modulators, e.g. inhibitors with the complex. For example, high resolution X-ray diffraction data collected from crystals exposed to different types of solvent allows the determination of where each type of solvent molecule resides. Small molecules that bind tightly to these sites can then be designed and synthesized and tested for their Aurora-A modulation, e.g. inhibition activity.

All of the complexes referred to above may be studied using well-known X-ray diffraction techniques and may be refined versus 1.5-3 A resolution X-ray data to an R value of about 0.20 or less using computer software, such as X-PLOR [Yale University, 1992, distributed by Molecular Simulations, Inc.; see, e. g., Blundell & Johnson, supra; Meth. Enzymol., vol. 114 & 115, H. W. Wyckoff et al., eds., Academic Press (1985)]. This information may thus be used to optimize Aurora-A agonists, partial agonists, antagonists, partial antagonists, and more importantly, to design new Aurora-A agonists/antagonists.

Accordingly, the present invention is also directed to a binding site in Aurora-A for an Aurora-A ligand in which a portion of Aurora-A ligand is in van der Walls contact or hydrogen bonding contact with at least one of the following residues: Y199, L178, W128, L159, H187, L188, I184 of Aurora-A.

For purposes of this invention, by Aurora-A binding site it is also meant to include mutants or homologues thereof. In a preferred embodiment, the mutants or homologues have at least 25% identity, more preferably 50% identity, more preferably 75% identity, and most preferably 95% identity to residues Q127, W128, R126, L159, F157, E170, L169, V206, Y199, H187, R179, L178, V182, Y199, L188, I184, V252, K250, P282, H280 of Aurora-A binding sites.

The present invention is also directed to a machine-readable data storage medium, comprising a data storage material encoded with machine readable data, wherein the data is defined by the structure coordinates of an Aurora-A ligand complex according to Table B or a homologue of said complex, wherein said homologue comprises backbone atoms that have a root mean square deviation from the backbone atoms of the complex of not more than 3.0 Å, preferably of not more that 2.0 Å.

The present invention also provides for computational methods using three-dimensional models of the Aurora-A kinase that are based on crystals of the kinase or the kinase ligand complex. Generally, the computational method of designing an Aurora-A ligand determines which amino acid or amino acids of Aurora-A interact with a chemical moiety (at lease one) of the ligand using a three-dimensional model of a crystallized protein comprising the Aurora-A kinase with a bound ligand and selecting a chemical modification (at least one) of the chemical moiety to produce a second chemical moiety with a structure that either decreases or increases an interaction between the interacting amino acid and the second chemical moiety compared to the interaction between the interacting amino acid and the corresponding chemical moiety on the natural ligand, i.e. TPX2 or a fragment thereof.

The computational methods of the present invention are for designing kinase synthetic ligands using such crystal and three dimensional structural information to generate synthetic ligands that modulate the conformational changes of the kinase. These computational methods are particularly useful in designing an agonist, partial agonist, antagonist or partial antagonist to the kinase, wherein the agonist, partial agonist, antagonist or partial antagonist has an extended moiety that prevents any one of a number of ligand-induced molecular events that alter the kinase's influence on one of its targets, such as preventing the normal coordination of the activation domain observed for a naturally occurring ligand or other ligands that mimic the naturally occurring ligand, such as an agonist. As described herein, synthetic ligands of the kinase will be useful in modulating kinase activity in a variety of medical conditions.

Aurora-A is known to comprise various segments as follows: 1) an amino-terminal ligand-binding site; and 2) an Aurora-A activation segement.

This modularity permits different segments and sites of each protein to separately accomplish different functions, although the segments and sites influence each other.

Exploiting the TPX2-binding site in the N-terminal lobe of Aurora-A has the advantage of blocking the activation of this oncogenic kinase and of blocking it specifically. An Aurora-A specific inhibitor would be an important cancer therapeutic agent in particular for breast and colon tumours.

The polypeptides referred to herein (e. g., Aurora-A, TPX2, and the like) may be produced by any well-known method, including synthetic methods, such as solid phase, liquid phase and combination solid phase/liquid phase syntheses; recombinant DNA methods, including cDNA cloning, optionally combined with site directed mutagenesis; and/or purification of the natural products, optionally combined with enzymatic cleavage methods to produce fragments of naturally occurring forms of Aurora-A and TPX2. The peptides can be expressed, crystallized, its three dimensional structure determined with a ligand bound (either using crystal data from the same kinase or a different kinase or a combination thereof), and computational methods used to design ligands to its ligand binding site, particularly ligands that contain an extension moiety that coordinates the activation segment of Aurora-A.

Typically Aurora-A and TPX2 as well as fragments thereof are purified to homogeneity for crystallization.

Purity of Aurora-A is measured with SDS-PAGE, mass spectrometry and hydrophobic HPLC. The purified Aurora-A for crystallization should be at least 97.5 % pure or 97.5%, preferably at least 99.0% pure or 99.0% pure, more preferably at least 99.5% pure or 99.5% pure.

Initially purification of the unliganded kinase can be obtained by conventional techniques, such as size exclusion chromatography, hydrophobic interaction chromatography (HPLC), ion exchange chromatography (HPLC), and heparin affinity chromatography.

To achieve higher purification for improved crystals of Aurora-A, it will be desirable to ligand shift purify the kinase using a column that separates the kinase according to charge, such as an ion exchange or hydrophobic interaction column, and then bind the eluted kinase with a

ligand, especially an agonist or partial agonist. The ligand induces a change in the kinase's surface charge such that when rechromatographed on the same column, the kinase then elutes at the position of the liganded kinase are removed by the original column run with the unliganded kinase. Usually saturating concentrations of ligand are used in the column and the protein can be preincubated with the ligand prior to passing it over the column.

More recently developed methods involve engineering a "tag" such as with histidine placed on the end of the protein, such as on the amino terminus, and then using a nickle chelation column for purification, Janknecht R., Proc. Natl. Acad. Sci. USA Vol 88: 8972-8976 (1991) incorporated by reference.

Typically, purified Aurora-A is equilibrated at a saturating concentration of ligand at a temperature that preserves the integrity of the protein. Ligand equilibration can be stablished between 2 and 37 °C, although the kinase tends to be more stable in the 2-20 °C range.

Preferably crystals are made with the hanging and/or sittiing drop methods.

Regulated temperature control is desirable to improve crystal stability and quality. Temperatures between 4 and 25 °C are generally used and it is often preferable to test crystallization over a range of temperatures. It is preferable to use crystallization temperatures from 18 to 25 °C, more preferably 20 to 23 °C, and most preferably 22 °C.

The Aurora-A ligand of this invention is any peptide, peptide mimetic or nonpeptide, including small organic molecules, that is capable of acting as a ligand for Aurora-A. In a preferred embodiment, the Aurora-A ligand is an Aurora-A modulator. By" Aurora-A modulator " it is meant an agonist or activator, a partial agonist or partial activator, an antagonist or inhibitor, or a partial antagonist or partial inhibitor of the Aurora-A kinase.

Agonists or partial agonists induce changes in kinases that place them in an active conformation that allows them to influence one of its targets. There may be several different ligand-induced changes in the kinase's conformation.

Antagonists or partial antagonists bind to kinases, but fail to induce conformational changes that alter the kinase's target-influencing properties or physiologically telcram conformations. Binding of an antagonist or partial antagonist can also block the binding and therefore the actions of an agonist or partial agonist.

Partial agonists, or partial antagonists, bind to kinases and induce only part of the changes in the kinases that are induced by agonists or antagonists, respectively. The differences can be qualitative or quantitative. Thus, a partial agonist or partial antagonist may induce some of the conformation changes induced by agonists or antagonists, respectively, but not others, or it may only induce certain changes to a limited extent.

As described herein, the unliganded kinase is in a configuration that is either inactive, has some activity or has phosphorylating activity. Binding of agonist ligands induces conformational changes in the kinase such that the kinase becomes more active and/or is protected from deactivation, in particular from dephosphorylation.

According to a particular embodiment, the present invention relates to allosteric inhibitors of Aurora-A. The binding of such an allosteric inhibitor to Aurora-A results in a conformational change of the kinase, thereby decreasing the kinases's activity, preferably by blocking its activation mechanism. The preferred binding site of an allosteric inhibitor according to the present invention is the TPX2-recognition groove where TPX2 binds Aurora-A, as described herein.

Consequently, an extended chemical moiety (or more) from the ligand that stabilizes the binding or contact of the TPX2 binding site with the binding site of Aurora-A can be designed. Typically such chemical moieties will extend past and away from the molecular recognition domain on the ligand and usually past the buried binding cavity of the ligand.

Ligand binding by the kinase is a dynamic process, which regulates kinase function by inducing an altered conformation.

The three-dimensional structure of the liganded Aurora-A kinase can be used in the development of new Aurora-A synthetic ligands. In addition, Aurora-A is overall well suited to modern methods including three-dimensional structure elucidation and combinatorial chemistry such as those disclosed in EP 335 628, U. S. patent 5,463,564, which are incorporated herein by reference. Computer programs that use crystallography data when practicing the present invention enable the rational design of ligands to Aurora-A. Programs such as RASMOL can be used with the atomic coordinates from crystals generated by practicing the invention or used to practice the invention by generating three dimensional models and/or determining the structures involved in ligand binding. Computer programs such as INSIGHT and GRASP allow for further manipulation and the ability to introduce new structures. In addition, high throughput binding and bioactivity assays can be devised using

purified recombinant protein and modern assays described herein and known in the art in order to refine the structure of a ligand and thereby the activity.

Generally the computational method of designing an Aurora-A synthetic ligand comprises two steps:

1) determining which amino acid or amino acids of Aurora-A interacts with a first chemical moiety (at least one) of the ligand using a three dimensional model of a crystallized protein comprising Aurora-A with a bound ligand; and
2) selecting chemical modifications (at least one) of the first chemical moiety to produce a second chemical moiety with a structure to either decrease or increase an interaction between the interacting amino acid and the second chemical moiety compared to the interaction between the interacting amino acid and the first chemical moiety.

Preferably the method is carried out wherein said three dimensional model is generated by comparing isomorphous ligand derivatives to produce improved phasing. Further preferred is wherein said method comprises determining a change in interaction between said interacting amino acid and said ligand after chemical modification of said first chemical moiety, especially wherein said three dimensional model is generated by comparing isomorphous ligand derivatives to produce improved phasing. Also preferred is wherein said selecting uses said first chemical moiety that interacts with at least one of the interacting amino acids Y199, L178, W128, L159, H187, L188, I184.

As shown herein, interacting amino acids form contacts with the ligand and the center of the atoms of the interacting amino acids are usually 2 to 4 angstroms away from the center of the atoms of the ligand. Generally these distances are determined by computer, however distances can be determined manually once the three dimensional model is made. See also Wagner et al., Nature 378 (6558): 670-697 (1995) for stereochemical figures of-three dimensional models. More commonly, the atoms of the ligand and the atoms of interacting amino acids are 3 to 4 angstroms apart. The invention can be practiced by repeating steps I and 2 to refine the fit of the ligand to the ligand bindiung site and to determine a better ligand, such as an agonist, partial agonist, antagonist or partial antagonist.

The three dimensional model of Aurora-A can be represented in two dimensions to determine which amino acids contact the ligand and to select a position on the ligand for chemical modification and changing the interaction with a particular amino acid compared to that before chemical modification. The chemical modification may be made using a

computer, manually using a two dimensional representation of the three dimensional model or by chemically synthesizing the ligand. The ligand can also interact with distant amino acids after chemical modification of the ligand to create a new ligand. Distant amino acids are generally not in contact with the ligand before chemical modification. A chemical modification can change the structure of the ligand to make a new ligand that interacts with a distant amino acid usually at least 4.5 angstroms away from the ligand, preferably wherein said first chemical moiety is 6 to 12 angstroms away from a distant amino acid. Often distant amino acids will not line the surface of the binding cavity for the ligand, they are too far away from the ligand to be part of a pocket or binding cavity. The interaction between an amino acid of the ligand binding site and an atom of a ligand can be made by any force or attraction described in nature. Usually the interaction between the atom of the amino acid and the ligand will be the result of a hydrogen bonding interaction, charge interaction, hydrophobic interaction, van der Waals interaction or dipole interaction. In the case of the hydrophobic interaction it is recognized that this is not a per se interaction between the amino acid and ligand, but rather the usual result, in part, of the repulsion of water or other hydrophilic group from a hydrophobic surface. Reducing or enhancing the interaction of the ligand binding site and a ligand can be measured by calculating or testing binding energies, computationally or using thermodynamic or kinetic methods as known in the art.

Chemical modifications will often enhance or reduce interactions of an atom of an amino acid of the ligand binding site and an atom of a ligand. Steric hindrance will be a common means of changing the interaction of the ligand binding cavity with the activation segment.

The present invention also provides methods for identifying compounds that modulate kinase activity. Various methods or combinations thereof can be used to identify these compounds. For example, test compounds can be modeled that fit spatially into the Aurora-A ligand binding site as defined by structure coordinates according to Table B, or using a three-dimensional structural model of Aurora-A, mutant Aurora-A or Aurora-A homolog or portion thereof. Structure coordinates of the ligand binding site, in particular amino acids Q127, W128, R126, L159, F157, E170, L169, V206, Y199, H187, R179, L178, V182, Y199, L188, I184, V252, K250, P282, H280 can also be used to identify structural and chemical features. Identified structural or chemical features can then be employed to design or select compounds as potential Aurora-A modulators. By structural and chemical features it is meant to include, but is not limited to, van der Waals interactions, hydrogen bonding interactions, charge interaction, hydrophobic bonding interaction, hydrophobic interaction and dipole interaction. Alternatively, or in conjunction, the three-dimensional structural model or the ligand binding site can be employed to design or select compounds as potential Aurora-A

modulators. Compounds identified as potential Aurora-A modulators can then be synthesized and screened in an assay characterized by binding of a test compound to the Aurora-A.

Examples of assays useful in screening of potential Aurora-A modulators include, but are not limited to, screening in silico, in vitro assays and high throughput assays, for instance, based on phosphorylation of histone H3 as described herein (Crosio, C., Fimia, G. M., Loury, R., Kimura, M., Okano, Y., Zhou, H., Sen, S., Allis, C. D., and Sassone-Corsi, P. (2002). Mitotic phosphorylation of Histone H3: spatio-temporal regulation by mammalian Aurora kinases. Mol. Cell Biol. 22, 874-885).

A preferred method of the invention can be described as a computational method of designing an kinase antagonist from an kinase agonist comprising:

1) determining a structure of a molecular recognition site of said agonist using a three dimensional model of a crystallized protein comprising an Aurora-A, and
2) selecting at least one chemical modification of said agonist that provides a ligand structure that extends beyond a binding site for said agonist and in the direction of at least one protein site important in Aurora-A biological function.

Another preferred method of the invention can be described as a computational method of designing a selective kinase modulator such as an kinase super agonist or antagonist comprising:

1) determining at least one interacting amino acid of an Aurora-A that interacts with at least one first chemical moiety of said ligand using a three dimensional model of a crystallized protein comprising Aurora-A with a bound ligand, and
2) selecting at least one chemical modification of said first chemical moiety to produce a second chemical moiety with a structure to reduce or enhance an interaction between said interacting amino acid and said second chemical moiety compared to said interaction between said interacting amino acid and said first chemical moiety.

However, as will be understood by those of skill in the art upon this disclosure, other structure based design methods can be used. Various computational structure based design methods have been disclosed in the art.

For example, a number computer modeling systems are available in which the sequence of the Aurora-A and the Aurora-A structure (i. e., atomic coordinates of Aurora-A and/or the atomic coordinates of the active site, the bond and dihedral angles, and distances between atoms in the active site such as provided in Table A and Table B) can be input. This

computer system then generates the structural details of the site in which a potential Aurora-A modulator binds so that complementary structural details of the potential modulators can be determined. Design in these modeling systems is generally based upon the compound being capable of physically and structurally associating with Aurora-A. In addition, the compound must be able to assume a conformation that allows it to associate with Aurora-A. Some modeling systems estimate the potential inhibitory or binding effect of a potential Aurora-A modulator prior to actual synthesis and testing.

Methods for screening chemical entities or fragments for their ability to associate with Aurora-A are also well known. Often these methods begin by visual inspection of the active site on the computer screen. Selected fragments or chemical entities are then positioned with the Aurora-A. Docking is accomplished using software such as QUANTA and SYBYL, following by energy minimization and molecular dynamics with standard molecular mechanic forcefields such as CHARMM and AMBER. Examples of computer programs which assist in the selection of chemical fragment or chemical entities useful in the present invention include, but are not limited to, GRID (Goodford, P. J. J. Med. Chem.1985 28: 849-857), AUTODOCK (Goodsell, D. S. and Olsen, A. J. Proteins, Structure, Functions, and Genetics 1990 8: 195-202), and DOCK (Kunts et al. J. Mol. Biol. 1982 161: 269-288).

Upon selection of preferred chemical entities or fragments, their relationship to each other and Aurora-A can be visualized and the entities or fragments can be assembled into a single potential modulator.

Programs useful in assembling the individual chemical entities include, but are not limited to CAVEAT (Bartlett et al. Molecular Recognition in Chemical and Biological Problems Special Publication, Royal Chem. Soc. 78,182-196 (1989)) and 3D Database systems (Martin, Y. C. J. Med. Chem. 1992 35: 2145-2154).

Alternatively, compounds may be designed de novo using either an empty active site or optionally including some portion of a known inhibitor. Methods of this type of design include, but are not limited to LUDI (Bohm H-J, J. Comp. Aid. Molec. Design 1992 6: 61-78) and LeapFrog (Tripos Associates, St. Louis. MO).

Examples of preferred ligands are found among the above-described indole and indene derivatives of formula (I), and optical isomers, physiologically acceptable salts, derivatives and prodrugs thereof.

These include especially indole and indene derivatives of formula (I), wherein

R^1 represents hydrogen, alkylene-NHR^8, alkylene-OR^8, or alkylene-SR^8;

R^2 represents hydrogen, alkylene-NHR^8, alkylene-OR^8, or alkylene-SR^8;

R^3 represents hydrogen, alkyl, alkylene-R^9, alkenylene-R^9, alkynylene-R^9, or arylene-R^9;

R^4 represents hydrogen;

R^5 represents hydrogen, alkyl, OR^{10}, NHR^{10}, SR^{10}, alkylene-R^{10}, alkenylene-R^{10}, alkynylene-R^{10}, or arylene-R^{10};

R^6 represents hydrogen, alkyl, OR^{10}, NHR^{10}, SR^{10}, alkylene-R^{10}, alkenylene-R^{10}, alkynylene-R^{10}, or arylene-R^{10};

R^7 represents hydrogen;

R^8 represents hydrogen, CO-alkyl, $(aa)_m asp(aa)_n$, $(aa)_m glu(aa)_n$, or $(aa)_m cys(aa)_n$;

R^9 represents NH-alkyl, $N(alkyl)_2$, $N^+(alkyl)_3$, aryl, or heteroaryl;

R^{10} represents hydrogen, aryl, or substituted aryl;

X represents a nitrogen atom or CH;

aa represents an amino acid residue; and

n is zero or an integer of 1 to 10;

m is zero or an integer of 1 to 10,

provided that R^1 and R^2 are not both hydrogen and that R^5 and R^6 are not both hydrogen, and optical isomers, physiologically acceptable salts and prodrugs thereof.

The physiologically acceptable salts in the present case can be acid addition or base addition salts.

For acid addition salts, inorganic acids, such as hydrochloric acid, sulphuric acid, nitric acid or phosphoric acid, or organic acids, in particular carboxylic acids, e.g. acetic acid, tartaric acid, lactic acid, citric acid, malic acid, mandelic acid, ascorbic acid, maleic acid, fumaric acid, gluconic acid or sulphonic acids, e.g. methanesulphonic acid, benzenesulphonic acid and toluenesulphonic acid, and the like are used.

The base addition salts include salts of the compounds of the formula (I) with inorganic bases, such as sodium hydroxide or potassium hydroxide, or with organic bases, such as mono-, di- or triethanolamine.

Prodrugs of the compounds of the formula I are, for example, physiologically easily hydrolysable esters such as alkyl, pivaloyloxymethyl, acetoxymethyl, phthalidyl, indenyl and methoxymethyl esters.

Derivatives of the compounds of the formula I are, for example, compounds wherein functional groups such as amino groups and carboxyl groups carry well-known protecting groups.

If the compounds according to the invention have asymmetric centres, racemates and optical isomers are included as mixtures or in pure form (enantiomers, diastereomers).

The term "alkyl" and the alkyl moieties of alkoxy, alkylthio and alkoxycarbonyl include straight-chain or branched alkyl groups, such as CH_3, C_2H_5, n-propyl, $CH(CH_3)_2$, n-butyl, $CH(CH_3)$-C_2H_5, isobutyl, $C(CH_3)_3$, n-pentyl or n-hexyl, in particular CH_3, C_2H_5 or $CH(CH_3)_2$, preferably having - if not stated otherwise - 1 to 8, in particular 1 to 6 and particularly preferred 1 to 4 carbon atoms.

The term "haloalkyl" and the haloalkyl moieties of haloalkoxy include straight-chain or branched alkyl groups having 1 to 4 carbon atoms, preferably 1 or 2 carbon atoms, where the hydrogen atoms in these groups may be partially or fully replaced by halogen atoms as mentioned above, for example C_1-C_2-haloalkyl, such as chloromethyl, bromomethyl, dichloromethyl, trichloromethyl, fluoromethyl, difluoromethyl, trifluoromethyl, chlorofluoromethyl, dichlorofluoromethyl, chlorodifluoromethyl, 1-chloroethyl, 1-bromoethyl, 1-fluoroethyl, 2-fluoroethyl, 2,2-difluoroethyl, 2,2,2-trifluoroethyl, 2-chloro-2-fluoroethyl, 2-chloro-2,2-difluoroethyl, 2,2-dichloro-2-fluoroethyl, 2,2,2-trichloroethyl and pentafluoroethyl. Most preferred are perhalogentated C_1-C_2-alkyl radicals, especially trifluormethyl and pentafluoroethyl.

The term "alkylene" includes straight-chain or branched alkylene groups, such as methylene and ethylene, preferably having - if not stated otherwise - 1 to 8, in particular 1 to 6 and particularly preferrred 1 to 4 carbon atoms.

The term "alkenylene" includes straight-chain or branched, mono- or polyunsaturated alkylene groups, such as ethenylene, preferably having - if not stated otherwise - 2 to 8, in particular 2 to 6 and particularly preferably 2 to 4, carbon atoms.

The term "alkynylene" includes straight-chain or branched, mono- or polyunsaturated alkylene groups, such as ethynylene, preferably having - if not stated otherwise - 2 to 8, in particular 2 to 6 and particularly preferably 2 to 4, carbon atoms.

The term "cycloalkyl" includes mono- or bicyclic hydrocarbon radicals, preferably having 5 to 14, in particular 5 to 8, carbon atoms and especially represents cyclopentyl, cyclohexyl, cycloheptyl and cyclooctyl.

The term "cycloalkenyl" includes mono- or bicyclic, unsaturated hydrocarbon radicals, preferably having 5 to 14, in particular 5 to 8, carbon atoms and especially represents cyclopentenyl, cyclohexenyl, cycloheptenyl and cyclooctenyl.

"Aryl" is in particular a mono- or bicyclic aromatic radical, preferably having 5 to 14, in particular 5 or 6, carbon atoms and especially represents naphthyl, indenyl, and in particular phenyl.

"Heteroaryl" is in particular a mono- or bicyclic heteroaromatic radical, preferably having 5 to 14, in particular 5 or 6, ring atoms and containing 1, 2 or 3 heteroatom(s) independently selected from the group consisting of O, N and S. According to a particular embodiment, bicyclic radicals contain a 5- or 6-membered heteroaromatic radical which is benzo-fused. Heteroaryl includes nitrogen-containing radicals, such as pyrrolyl, imidazolyl, pyrazolyl, pyridazinyl, pyrazinyl, indolyl, quinolinyl, especially pyridyl, pyrimidyl and isoquinolinyl, further nitrogen-containing radicals such triazolyl and tetrazolyl; radicals which contain an oxygen atom or a sulphur atom, such as thienyl, benzothienyl, furanyl and especially benzofuranyl; radicals which contain two or more different heteroatoms, such as thiazolyl, isothiazolyl, thiadiazolyl, isoxazolyl and oxazolyl. Preferred aromatic heterocyclic radicals are, pyridyl and indolyl.

"Saturated and partially- preferably mono - unsaturated heterocyclic radicals" in particular include 5- to 7-membered heteroaliphatic rings which containing 1, 2 or 3 heteroatom(s) independently selected from the group consisting of O, N and S. Examples include 3-pyrazolidinyl, 4-pyrazolidinyl, 5-pyrazolidinyl, 2-pyrrolidin-2-yl, 2-pyrrolidin-3-yl, 3-pyrrolidin-2-yl, 3-pyrrolidin-3-yl, 1-piperidinyl, 2-piperidinyl, 3-piperidinyl, 4-piperidinyl, pyridin(1,2-dihydro)-2-on-1-yl, 2-piperazinyl, morpholin-4-yl and thiomorpholin-4-yl.

Substituted aryl or heteroaryl radicals in particular include those having 1, 2 or 3 substitents independently selected from the group consisting of halogen, cyano, nitro, hydroxy, C_1-C_4-alkyl, C_1-C_4-alkoxy, C_1-C_4-alkylthio, C_1-C_4-haloalkyl, C_1-C_4-haloalkoxy and C_1-C_4-alkoxycarbonyl.

The term "arylene" is preferably naphthylene and in particular phenylene, such as 1,4-phenylene.

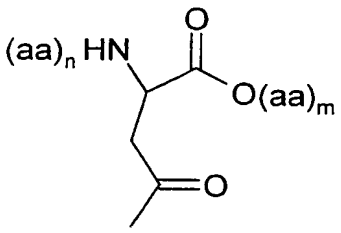
According to a preferred embodiment, one of residues R^1 and R^2, preferably residue R^2, is hydrogen while the other, preferably R^1, has a meaning different from hydrogen. In this context, alkylene radicals preferably contain a relatively short main chain connecting the indole or indene moiety with the amino, oxy or thio moiety (-NHR^8, -OR^8 or -SR^8, respectively). Accordingly, the main chain of preferred alkylene radicals contain 1 to 3 carbon atoms. Said alkylene radicals may be branched. Linear alkylene radicals are preferred. Specific examples of alkylene radicals include methylene, eth-1,1-ylene, prop-1,1-ylene and prop-2,2-ylene, with methylene being preferred.

According to a particularly preferred embodiment, one of residues R^1 and R^2, preferably R^2, is hydrogen and the other, preferably R^1, represents alkylene-NHR^8.

According to a further particularly preferred embodiment, one of residues R^1 and R^2, preferably R^2, is hydrogen and the other, preferably R^1, represents alkylene-OR^8.

R^8 preferably represents an acyl radical so as to form an amide, ester or thioester bond within residue R^1 and/or R^2. It is preferred that residue R^8 represents an amino acid selected from the group consisting of aspartate (asp), glutamate (glu), and cysteine (cys), i.e., n and m are zero, or a peptide connected via aspartate, glutamate or cysteine, i.e., aa represents an amino acid radical, and at least one of n and m is an integer of 1 or higher, wherein the amino acid radicals can be the same or different. The aspartate, glutamate and cysteine radicals may be bonded to the alkylene radical via any of their functional groups. The aspartate and glutamate radicals are preferably bonded via any of its carboxy groups, the β-carboxy function of aspartate and the γ-carboxy function of glutamate being preferred. Cystein radicals are preferably bonded via their side chain thio functionality. Accordingly, the amino acid radicals aa are connected to the alkylene group via the remaining functional groups of the aspartate, glutamate or cysteine radicals. According to a preferred embodiment, the remaining functional groups are the α-amino group and the α-carboxy group, of which one or both can be bonded to the amino acid radicals aa. According to this embodiment, the connecting aspartate, glutamate or cystein radical is itself part of the peptide chain which is connected to the alkylene residue via the side chain of said aspartate, glutamate or cystein radical. In case both of n and m are zero, the remaining functional groups of the aspartate, glutamate or cysteine radicals may be derivatized, e.g. with groups that are well-known as corresponding protecting groups from petide synthesis. These include

amide-forming acyl groups such as an acetyl group for the α-amino group or ester-forming alkyl groups such as an methyl group for the α-carboxy group. Analoguously, the remaining functional groups of any amino acid aa (in case m and/or n is not zero) can be derivatized in a similar manner. A specific example of such a preferred residue R^8 is a radical of the formula (II)



(II)

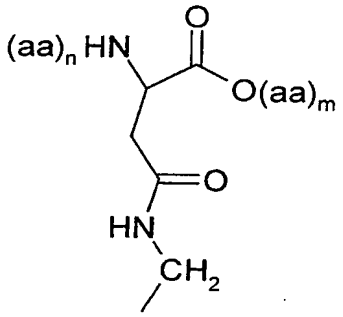
wherein

aa represents an amino acid radical;

n is zero or an integer of 1 to 10; and

m is zero or an integer of 1 to 10.

A specific example of such preferred residues R1 and/or R2 is thus the radical of formula (III)



(III)

wherein

aa represents an amino acid radical;

n is zero or an integer of 1 to 10; and

m is zero or an integer of 1 to 10.

The amino acid radicals aa may be any amino acid. Preferably said amino acid radicals are selected to mimick – together with the connecting amino acid - a portion of TPX2 that binds to Aurora-A. Said portion of TPX2 is in particular a portion comprising the amino acid sequence from residues 7 to 21, preferably 8 to 19 (compare Fig. 3C) or a fragment thereof. Accordingly, suitable peptides may have the sequence Tyr-Xaa1-Tyr-Xaa2-Ala-Pro-Xaa3-Xaa4-Phe-Xaa5-Xaa6-Phe or a portion thereof, wherein Xaa1-6 may be any amino acid radical. Xaa1 preferably is a serine radical or a similar amino acid. Xaa2 preferably is an aspartate radical or a similar amino acid. Xaa3 preferably is a serine radical or a similar amino acid. Xaa4 preferably is an aspartate radical or a similar amino acid. Xaa5 preferably

is an isoleucine radical or a similar amino acid. Xaa6 preferably is an asparagine radical or a similar amino acid. A particularly preferred portion is the sequence Tyr-Xaa1-Tyr-Xaa2-Ala-Pro-Xaa3-Xaa4-Phe. By similar amino acid is meant an amino acid that is considered to result in a conservative change of the peptide's structure when it replaces the amino acid to which it is similar. For instance, aspartate is similar to glutamate. The peptide preferably has a lenght of 3 to 15, more preferably of 4 to 10 and advantagously of 5 to 8 amino acid, the sum of n + m thus being 2 to 14, more preferably of 3 to 9 and advantagously of 4 to 7, respectively.

According to a further preferred embodiment, R^8 represents hydrogen, especially in case R^2 is alkylene-OR^8 or alkylene-SR^8.

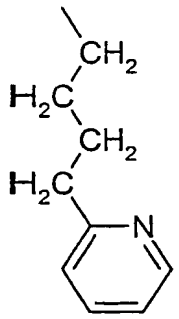
According to a further embodiment, R^8 represents alkyl, aryl or heteroaryl, especially in case R^2 is alkylene-NHR^8.

According to a further embodiment, one of residues R^1 and R^2, preferably R^2, is hydrogen and the other, preferably R^1, represents hydrogen, C_1-C_4-alkyl (preferably methyl) or C_1-C_4-haloalkyl (preferably trifluormethyl).

R^3 preferably is a radical different from hydrogen and advantageously represents alkylene-R^9, alkenylene-R^9, or alkynylene-R^9. In this context, the main chain connecting the indene or indole moiety with residue R^9 preferably is relatively long. Accordingly, it is preferred that the main chain of the alkylene radical, alkenylene radical or alkynylene radical contains 3 to 8, e.g. 4, carbon atoms. Said alkylene radical, alkenylene radical and alkynylene radical may be branched. Linear radicals are preferred. Specific examples of said radicals include prop-1,3-ylene, but-1,4-ylene, pent-1,5-ylene and hex-1,6-ylene, with but-1,4-ylene being preferred.

R^9 preferably represents aryl or heteroaryl. Specific examples of aryl and heteroaryl radicals include phenyl, naphthyl, indenyl, pyridyl, and indolyl, with pyridyl, e.g. pyrid-2-yl, being preferred. Further specific examples of aryl and heteroaryl radicals include imidazolyl, triazolyl and tetrazolyl. R^9 may also represent substituted aryl or heteroaryl.

A specific example of such a preferred residue R^3 is a radical of the formula (IV)



(IV)

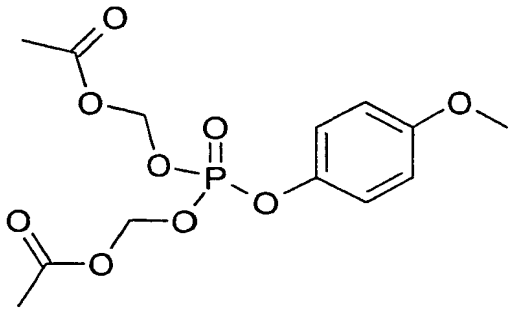
While residues R^4, R^5, R^6 and R^7 all may be hydrogen, it is preferred that at least one of said residues, especially R^6, is different from hydrogen.

R^5 and/or R^6 preferably represent OR^{10}, NHR^{10}, SR^{10}, or alkylene-R^{10}. In this context, the preferred embodiments regarding the alkylene radical are those as described in connection with residues R^1 and R^2.

R^{10} preferably represents aryl or substituted aryl. In this context, aryl is preferably phenyl. Substituted aryl, in general, contains 1, 2 or 3 substituents which may be the same or different, mono-substitution being preferred. The substituents are preferably selected from the group consisting of hydroxy, $-OPO_3H_2$, $-CH_2PO_3H_2$, $-CF_2PO_3H_2$, -COOH, $-CH(COOH)_2$, $-OPO_3(R^{11})_2$, $-CH_2OPO_3(R^{11})_2$, $-CF_2PO_3(R^{11})_2$, $-COOR^{11}$, and $-CH(COOR^{11})_2$, wherein R^{11} is a radical that is cleavable *in vivo*, converting the carboxylic acid esters, phosphate and phosphonate esters to carboxylates, phosphates or phosphonates, respectively. Suitable examples of R^{11} residues are alkyl, CH_2OCO-alkyl, and C_2H_4-S-CO-alkyl. Especially preferred substituents are selected from the group consisting of $-OPO_3H_2$, $-CH_2PO_3H_2$, $-CF_2PO_3H_2$, $-OPO_3(R^{11})_2$, $-CH_2OPO_3(R^{11})_2$, and $-CF_2PO_3(R^{11})_2$. A preferred example of a R^{11} residue is $-CH_2OCO$-alkyl, alkyl being linear or branched and having 1 to 6 carbon atoms, e.g. methyl. Alternatively, hydroxy is an especially preferred substituent. Further substituents include halogen, cyano, nitro, C_1-C_4-alkyl, C_1-C_4-alkoxy, C_1-C_4-alkylthio, C_1-C_4-alkoxycarbonyl.

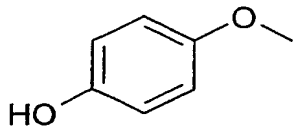
R^{10} may also be a moiety that is similar to the above-described aryl or substituted aryl radicals. For instance, R^{10} may be a cyclic moiety such as a saturated, partially unsaturated or aromatic, alicyclic or heterocyclic radical which may be substituted. These cyclic moieties include in particular heteroaryl, cycloalkyl, cycloalkylenyl, saturated heterocyclic and mono-unsaturated heterocyclic radicals, wherein the heterocyclic radicals preferably are 5- to 7-membered and contain 1 or 2 heteroatoms selected from the group consisting of O, N and S. The substituents preferably include those for the aryl radicals.

A specific example of such preferred residues R5 and/or R6 is the radical of formula (V)



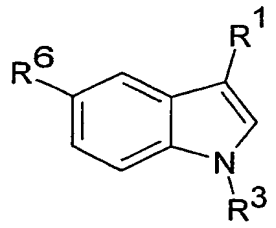
(V)

A further specific example of such preferred residues R5 and/or R6 is the radical of formula (VI)



(VI)

According to a specific embodiment, the present invention relates to indole derivatives of formula (Ia)



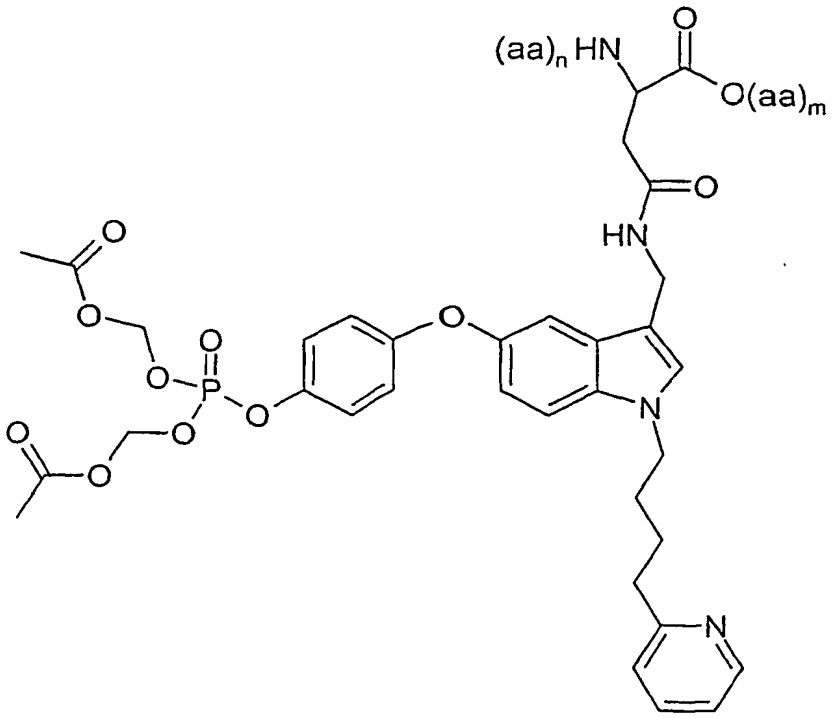
(Ia)

wherein

R^1, R^3 and R^6 are defined as above,

and optical isomers, physiologically acceptable salts and prodrugs thereof.

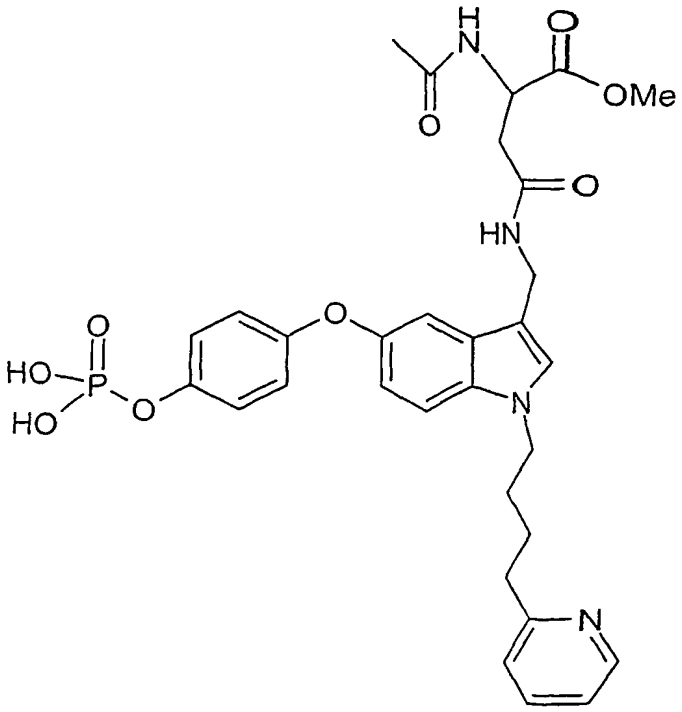
A preferred indole derivative of the present invention has the formula (15)



15

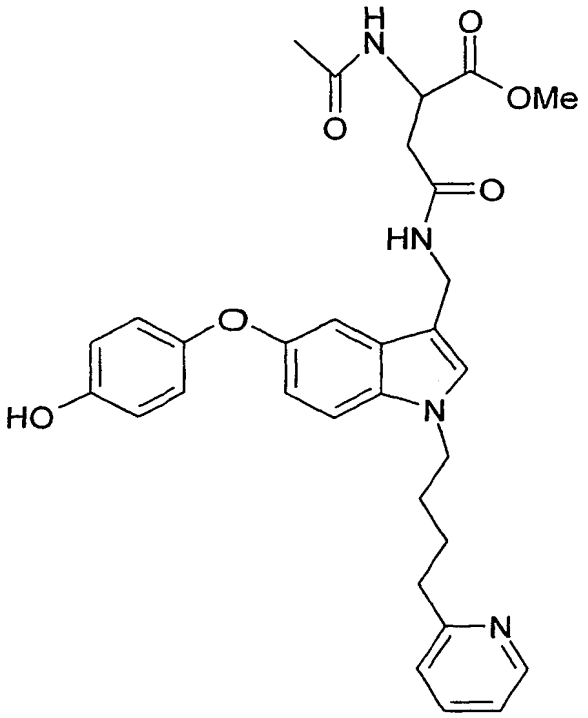
wherein
aa, n and m are defined as above.

A further preferred indole derivative of the present invention has the formula (14)



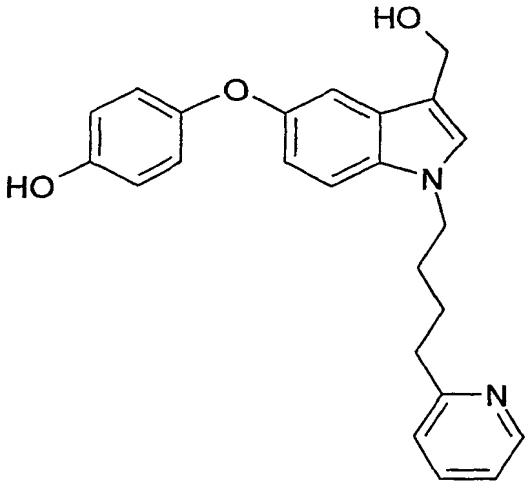
14

A further preferred indole derivative of the present invention has the formula (12)



12

A further preferred indole derivative of the present invention has the formula (7)



7

The indene and indole derivatives of formula (I) can be prepared by methods well-known to those skilled in the art.

For instance, the above 3-hydroxymethylindole compound (7) is obtainable by (i) alkylating the NH group of 5-benzyloxyindole with 1,3-dibromopropane; (ii) removing the benzyl protecting group by catalytic hydrogenolysis; (iii) reacting the hydroxy group with a benzyl-

protected p-hydroxyphenylboronic acid; (iv) reacting the 3'-bromine with 2-methylpyridine; (v) introducing an aldehyde group in 3-position of the indole nucleus; and (vi) converting the aldehyde into a hydroxymethyl group by hydrogenolysis.

Further, the above 3-N-aspartylaminomethyl-5-O-*p*-hydroxyphenyl-1-N-(4-pyridylbutyl)-indole (12) is obtainable by (i) reacting the aldehyde from step (v) above with hydroxylamine to the corresponding oxime; (ii) reducing the oxime; (iii) reacting the aminomethyl group with N-acetylaspartyl methylester; and (iv) removing the benzyl-protecting group by catalytic hydrogenolysis.

Further, the above 3-N-aspartylaminomethyl 5-O-*p*-hydroxyphenyl-1-N-(4-pyridylbutyl)-indole phosphate (14) is obtainable by phosphorylating compound (12) using dibenzyl-N,N-diisopropylaminophosphoramidite and subsequent catalytic hydrogenolysis to remove the benzyl-protecting group.

Moreover, further amino acids can be readily attached to the N,O-protected aspartyl group using well-known peptide chemistry. To this end, the amino- and/or carboxy-protecting groups can be removed in a manner know per se and reacted with appropriate amino acid derivatives which usually are themselves protected as required. Both solid phase and solution synthesis can be employed.

The present invention is further directed to a method for treating cancer, in particular breast and colon carcinomas, comprising administering an effective amount of an Aurora-A modulator, such as the indole and indene derivatives as described herein, preferably an antagonist or partial antagonist, identified by a computational process of the invention.

The compounds according to the invention are thus suitable for the treatment of disorders in which the interaction of Aurora-A with TPX2 is responsible for the formation or the progressive course of these disorders. In particular, the compounds according to the invention can be used for the treatment of cancer.

The compounds according to the invention can either be administered as individual therapeutic active compounds or as mixtures with other therapeutic active compounds: they can be administered as such, but in general they are administered in the form of pharmaceutical compositions, i.e. as mixtures of the active compounds with pharmaceutically acceptable excipients, in particular vehicles or diluents and/or additives. The compounds or

compositions can be administered enterally, e.g. orally or rectally, or parenterally, e.g. subcutaneously, intravenously or intramuscularly.

The nature of the pharmaceutical composition and of the pharmaceutical carrier or diluent depends on the desired manner of administration. Oral compositions can be present, for example, as tablets or capsules and can contain customary excipients, such as binding agents (e.g. syrup, acacia, gelatin, sorbitol, tragacanth or polyvinylpyrrolidone), fillers (e.g. lactose, sugar, maize starch, calcium phosphate, sorbitol or glycine), lubricants (e.g. magnesium stearate, talc, polyethylene glycol or silica), disintegrating agents (e.g. starch) or wetting agents (e.g. sodium laurylsulphate). Oral liquid preparations can be present in the form of aqueous or oily suspensions, solutions, emulsions, syrups, elixirs or sprays etc. or can be present as dry powders for reconstitution with water or another suitable carrier. Liquid preparations of this type can contain customary additives, for example suspending agents, flavourings, diluents or emulsifiers. For parenteral administration, solutions or suspensions with customary pharmaceutical carriers can be employed.

The following examples are to illustrate the invention, but should not be interpreted as a limitation thereon.

Examples

Experimental Procedures

Protein and Antibody Preparation

AuroraΔN was cloned in pET M11 (residues 122-403, wild-type and D274N) were expressed in CodonPlus RIL *E.coli* (Stratagene) and purified with TALON resin (Clontech) using manufacturer's instructions. AuroraΔN was treated with TEV protease to remove the His tag, passed through a TALON column to remove TEV and then purified to homogeneity by size exclusion chromatography. Human and *Xenopus* GST TPX2 fragments were expressed in BL21(DE3) *E.coli*. TPX2 proteins were purified using glutathione sepharose (Pharmacia). To form the AuroraΔN-TPX2 complex, cell lysate from *E.coli* expressing GST TPX2 was passed through a glutathione sepharose column, washed, and then purified AuroraΔN was passed through the same column. Column resin was incubated with TEV protease to elute the complex, which was further purified by size exclusion chromatography.

The polyclonal antibody against full-length TPX2 was produced in rabbits with bacterially expressed GST TPX2 (Wittmann et al., 2000) and affinity purified. The 1C1 monoclonal anti-*Xenopus* Aurora-A is described in: Giet R, Uzbekov R, Cubizolles F, Le Guellec K, Prigent C., J Biol Chem. (1999) 274, 15005-13. The polyclonal anti-human Aurora-A was obtained by injecting rabbits with bacterially expressed full-length human Aurora-A.

HeLa cell extract and Xenopus egg extract preparation

HeLa cells (CCL2; ATCC, Manassas, VA) were grown in 10% fetal calf serum (FCS) and 2 mM L-glutamine in modified Eagle's medium at 37°C and 5% CO_2. Cells were arrested in S phase using a double thymidine (2 mM) block as described (Stein, 1994) and released from S phase by washing away the thymidine. After 10 h, cells were trypsinized, incubated for 10 min in ice-cold lysis buffer (25 mM Tris pH 7.6, 200 mM NaCl, 1% Triton, protease inhibitors), and centrifuged for 10 min at 15,000 g. *Xenopus* cytostatic factor arrested extracts (CSF extracts) were prepared as described (Murray, 1991).

***In vitro* kinase assay**

Full-length TPX2, TPX2(1-43) or TPX2(15-43) at 4 μM were incubated with full length Aurora-A or AuroraΔN in kinase buffer (20 mM Hepes pH 7.5, 200 mM KCl, 5 mM MgCl2, 0.5 mM EGTA, 1 mM DTT, 0.05 % Triton X-100, 50 μM ATP) containing γ^{32}P-ATP (Amersham) in the presence or absence of histone H3 (0.2 mg/ml, Roche), for 15 min at 25°C. For cleavage, GST TPX2(1-43) was incubated with TEV 16-18h at 4° C (Figure 2C) or by adding TEV after the kinase reaction and incubating 10 min at 30° C (Figure 1D). In Figure 2D, AuroraΔN (8μM) was incubated in kinase buffer containing histone H3 and γ^{32}P-ATP for 15

min at 25° C. After separation by SDS-PAGE, phosphorylated histone H3 and TPX2 proteins were detected by autoradiography.

Phosphatase assay

Active human Aurora-A (2 μM) was incubated with TPX2, TPX2(1-43) or TPX2(15-43) (3 μM) for 10 min at 4°C. Samples were diluted in phosphatase buffer (50 mM Tris-HCl pH 7, 0.1 mM EDTA, 1 mM $MnCl_2$, 5 mM caffeine, 5 mM DTT, 0.025% Tween 20) and incubated in the absence or presence of PP1 (α isoform, NEB) for 0.5h at 30°C. Afterwards proteins were separated by SDS-PAGE and analyzed by Western blotting with anti-Aurora-A or anti-phospho-Aurora-A (Cell Signalling, 1:10000 dilution).

Immunoprecipitation

For coating beads with anti-GST, 6 μg of antibody were incubated with 20 μl of protein A-conjugated Dynabeads 280 (Dynal) in a total volume of 100 μl PBS-T (PBS, 0.1 % Triton X-100), for at least 1 h at 4°C. The beads were washed twice with PBS-T and twice with HeLa cells lysis buffer or CSF-XB (10 mM Hepes pH 7.7, 50 mM sucrose, 100 mM KCl, 2 mM $MgCl_2$, 0.1 mM $CaCl_2$ and 5 mM EGTA). 50 μl extract was incubated with GST, GST TPX2 1-43, GST TPX2 15-43, GST TPX2 *Xenopus* or GST TPX2(1-39) *Xenopus* proteins (final concentrations: 250 nM *Xenopus*; 1 μM HeLa) in the absence or presence of RanQ69L-GTP (16 μm) for 15 minutes at 20 °C and then added to the beads. After 1 h incubation on ice, the beads were washed twice with lysis buffer or CSF-XB, washed twice with PBS-T and boiled in SDS-PAGE sample buffer. The samples were then subjected to gel electrophoresis and analyzed by Western blotting with anti-GST, anti-*Xenopus* TPX2, anti-*Xenopus* Aurora-A or anti-human Aurora-A.

In vitro pull-down assays

GST or GST TPX2(1-43) (240 μg) was bound to 100 μl glutathione sepharose (Pharmacia) in 1ml binding buffer (PBS, 0.05% Tween-20, 2 mM DTT) for 0.5 h. Beads were washed three times, divided into four equal quantities in 1 ml binding buffer and 35 μg full-length human Aurora-A, AuroraΔN or AuroraΔN(D274N) was added. After binding for 0.5 h at 4°C, the resin was washed four times, then incubated overnight in 50 μl binding buffer plus 2 μg TEV protease. 10 μl supernatant from each was analysed by SDS-PAGE. TEV cleavage was necessary because AuroraΔN and GST TPX 1-43 co-migrate by SDS-PAGE.

Crystallographic methods

Needle-like crystals of AuroraΔN in complex with TPX2 1-43, approximately 10 μm in thickness, were grown at 18°C by vapour diffusion using 18% (w/v) PEG8000, 100 mM MES pH 6.5, 200 mM $MgSO_4$ as well buffer and hanging drops comprising a 1:1 mix of 20 mg/ml complex pre-mixed with 2 mM ATPγS and 0.2 mM $MgSO_4$, and 20% PEG8000, 100 mM MES pH 6.5, 200 mM $MgSO_4$. For cryo-protection, the drop buffer was supplemented with 17.5% (v/v) glycerol. Hexagonal prism crystals of the wild-type AuroraΔN kinase alone, approximately 50μm in all dimensions, were grown by vapour diffusion using 20% PEG300, 5% PEG8000, 100 mM Tris 8.5, 10% glycerol as well buffer and sitting drops comprising a 1:1 mix of well buffer and 9 mg/ml protein pre-incubated with 2 mM ATPγS and 2 mM $MgSO_4$. Diffraction data were collected at 100K and processed using the CCP4 suite of programs (CCP4, 1994). Structures were solved by molecular replacement using CNS (Brunger et al., 1998) and the coordinates of the cAMP-dependent protein kinase catalytic subunit as an initial model (Mashhoon et al., 2001). CNS was used for refinement and O (Jones et al., 1991) for model building. The statistics for data collection and refinement are shown in Table 1. Structure figures were prepared using PyMOL (DeLano, 2002). Coordinates and structure factors have been deposited in the Protein Data Bank with accession codes 1ol7 and 1ol5 for phosphoryated AuroraΔN alone and in complex with TPX2(1-43), respectively.

Mapping of a minimal domain of TPX2 sufficient for binding and activating Aurora-A

A minimal domain of TPX2 sufficient to interact with Aurora-A was mapped. The Aurora-A binding domain of TPX2 resides within its N-terminal 150 residues. Inspection of a sequence alignment of TPX2 shows that only a third of this region is conserved across species (residues 1-43 in human TPX2, 1-39 in *Xenopus* TPX2). This fragment is able to co-immunoprecipitate Aurora-A from *Xenopus* egg extracts to the same extent as the full-length protein (Figure 1A, lanes 6 and 4 respectively). In the egg extracts, efficient binding of TPX2 to Aurora-A requires Ran locked into its GTP-bound form (RanQ69L), which releases TPX2 from a complex with importin α/β. *Xenopus* TPX2(1-39) binds Aurora-A in a RanGTP-independent manner (Figure 1A, lanes 5 and 6), consistent with it lacking the portion of TPX2 regulated by Ran. The corresponding fragment of human TPX2 (residues 1-43) is able to interact with Aurora-A using pull-down experiments in HeLa extract, while a shorter construct encompassing residues 15-43 is unable to do so (Figure 1B, lanes 2 and 3 respectively).

The minimal Aurora-A binding domain of human TPX2(1-43) is sufficient to stimulate the activity of the kinase *in vitro* (Figure 1C). Phosphorylation of histone H3, an Aurora-A substrate, is markedly increased by the addition of TPX2(1-43) (Figure 1C compare lane 1 with lane 3), but not by the shorter TPX2(15-43) (lane 4). Full-length TPX2 also increases the

activity of the kinase, but appears less effective than the fragment. Full-length TPX2 may require additional factors such as microtubules for full activation. While microtubules can stimulate TPX2 activation of Aurora-A, no enhancement is detected for TPX2 1-43, which does not bind microtubules. Full-length TPX2 is itself a substrate for Aurora-A. The phosphorylation observed for GST-fused TPX2(1-43) (Figure 1C, lane 3) is however likely a non-specific artifact of the *in vitro* reaction because upon cleavage of the fusion protein phosphorylation signal is also detected on GST (lane 2). Furthermore, mass spectroscopy does not reveal Aurora-A phosphorylation sites in full-length TPX2.

The presence of TPX2(1-43) protects Aurora-A from deactivating dephosphorylation (Figure 1E). In the absence of TPX2, Aurora-A is completely dephosphorylated by phosphatase PP1 (lanes 1 and 2). In the presence of full-length TPX2, PP1 treatment dephosphorylates most sites of Aurora-A (lanes 3 and 4, upper panel) but prevents dephosphorylation of $Thr288^{AUR}$ (lane 4, lower panel, arrow). An even stronger protection effect is observed upon addition of TPX2(1-43) (lane 6). In the presence of TPX2 15-43, Aurora-A is fully dephosphorylated as in the absence of TPX2 altogether (lanes 8 and 2), consistent with the shorter fragment being unable to bind Aurora-A (Figure 1B). Thus, residues 1-43 of human TPX2 are necessary and sufficient for Aurora-A binding, activation and protection from dephosphorylation.

Overall structure of TPX2-bound Aurora-A

For crystallisation purposes, a minimal functional complex of Aurora-A and TPX2 was focused on. The N-terminus of Aurora-A is overall poorly conserved across species and is not required to bind TPX2. A fragment of Aurora-A lacking the N-terminus (residues 122-403, AuroraΔN) interacts with the active fragment of TPX2 (1-43) as efficiently as the full-length kinase (Figure 2A lanes 5 and 4). Both fragments are highly conserved (Figure 2D, 2E) and indeed *Xenopus* TPX2(1-39) can bind human AuroraΔN. AuroraΔN is phosphorylated when expressed in *E. coli*, as detected by a phospho-specific antibody (Figure 2B, lane 1). This appears to be due to Aurora-A autophosphorylation rather than to the activity of a bacterial kinase, since no phosphorylation can be detected upon expressing a mutant where an important catalytic residue, $Asp274^{AUR}$, has been mutated to Asn (Figure 2B, lane 2). The unphosphorylated mutant is catalytically inactive but retains TPX2-binding activity (Figure 2A, lane 6). AuroraΔN is activated by TPX2(1-43) (Figure 2C, lanes 2,3).

The complex between human phosphorylated AuroraΔN and TPX2(1-43) is active in histone H3 phosphorylation (Figure 2C, lane 7). The complex was crystallized in the presence of Mg^{2+} ions and the ATP analogue ATPγS. The structure was determined by molecular replacement and refined it using 2.5 Å resolution data to a *R*free of 25.2% (see Table 1). The

polypeptide model includes the catalytic core of the kinase (residues 123-387^{AUR}) and two segments of TPX2 (residues 7-21 ^{TPX} and 30-43^{TPX}). No ordered electron density is present for the eight intervening residues of TPX2 (22-29^{TPX}), which have therefore not been modelled.

The kinase catalytic core has an overall bilobate structure (Figure 3A,B). Briefly, the N-terminal lobe (residues 123-210) consists of a β-sheet and two α-helices, including the prominent helix αC whereas the C-terminal and larger lobe (residues 217-387) is mostly α-helical. The active site is situated at the interface between the lobes and includes the ATP-binding site, the catalytic base (Asp256^{AUR}) and the kinase activation segment (residues 274-299^{AUR}). In contrast to the structure of unphosphorylated Aurora-A, the activation segment is well ordered in the electron density and includes two phosphorylated threonine residues (Thr287^{AUR} and Thr288^{AUR}). Although the crystals were formed in the presence of ATP-γS, the electron density for the nucleotide reveals only the adenosine with two ordered phosphates and has been modelled as an ADP.

Specific recognition of TPX2 at two sites on Aurora-A

TPX2 binds Aurora-A with two separate stretches recognized at two distinct sites on the kinase. The upstream stretch (residues 7-21^{TPX}) binds at the N-terminal lobe of Aurora-A (Figure 3A,B). The downstream stretch (residues 30-43^{TPX}) binds in a α-helical conformation between the N- and C-terminal lobes (Figure 3A,B). The two Aurora-A-binding motifs of TPX2 appear to be connected by a flexible linker (disordered in the structure) that is variable in length and sequence across species (Figure 2D). Additionally, the two stretches are connected by an intramolecular hydrogen bond between Asp11^{TPX} and Trp34^{TPX}.

The upstream stretch of TPX2 has a mostly extended conformation, with a kink in the middle induced at a proline residue (Pro13^{TPX}) (for details see Figure 3C). The conserved segment ^{8}YSYDAPS^{14} (Figure 2D) is engaged in extensive main-chain and side-chain interactions with Aurora-A. In particular, Tyr8^{TPX}, Tyr10^{TPX} and Ala12^{TPX} tightly nestle into a hydrophobic groove between the β-sheet, helix αB and helix αC. An adjacent hydrophobic groove accommodates the side chains of TPX2 residues from Phe16^{TPX} to Phe19^{TPX}. The N-terminal residues of the Aurora-A catalytic core make key contributions to this interface, in particular with Arg126^{AUR} forming a cation-π interaction with Phe16^{TPX}.

The downstream helical stretch of TPX2 interacts with both helix αC and the activation segment of Aurora-A, bridging them (see Figure 3D for details). Most prominently, two conserved aromatic residues (Trp34^{TPX} and Phe35^{TPX}) interact with His187^{AUR} and

His280^{AUR}, and Ala39^{TPX} additionally contacts the activation segment at Pro282^{AUR}. While the side of the TPX2 helix in contact with the kinase is lined by hydrophobic and conserved residues, the opposite side exposed to solvent comprises hydrophilic and variable residues. Residues 40-43^{TPX} at the end of the helix assume an extended conformation and are involved in contacts with a symmetry-related Aurora-A molecule within the crystal lattice.

TPX2 is phosphorylated in mitotic extracts and may well be regulated by phosphorylation. Even if TPX2 was phosphorylated in the first 43 residues, examination of the crystal structure suggests that the interaction of TPX2 with Aurora-A is unlikely to be regulated this way because all the conserved serine residues point towards solvent.

Phosphorylated Aurora-A is in an active conformation when bound to TPX2

Comparison with other kinase structures reveals that phosphorylated TPX2-bound Aurora-A closely matches the conformation of kinases in the active conformation. Using the program DALI (Holm and Sander, 1993), it was found particularly similar to cAMP-dependent protein kinase (PDB code 1YDS, 1.3 Å r.m.s.d., 257 structurally equivalent residues). Structural comparison shows that all the conserved residues at the active site are correctly oriented for catalysis (Figure 4A). These include the positively-charged residue aligning the phosphates for catalysis (Lys162^{AUR}, equivalent to Lys72^{cAPK}), the negatively-charged residue coordinating the magnesium ion bound to the nucleotide (Asp274^{AUR}, equivalent to Asp184^{cAPK}), and the catalytic base (Asp256^{AUR}, equivalent to Asp166^{cAPK}) whose role is to transfer the γ-phosphate of ATP to the hydroxyl group of a substrate serine or threonine residue.

The phosphorylated activation segment of TPX2-bound Aurora-A is in a conformation typical of active Ser/Thr kinases. It is virtually superposable to that of active cAPK (Figure 4A), with an r.m.s.d. of 0.8 Å between the Cαatoms of residues 274-299^{AUR} and 184-208^{cAPK}. In cAPK, the phosphorylated Thr197^{cAPK} interacts with Arg165^{cAPK} and with a basic surface patch (His87^{cAPK} and Lys189^{cAPK}), influencing the preceding catalytic residue (Asp166^{cAPK}). In Aurora-A, the phosphoryl moiety of Thr288^{AUR} is at the same structural position and interacts with the corresponding Arg255^{AUR} and with a similar, though not identical, basic patch (Arg180^{AUR} and Arg286^{AUR}), linking the catalytic residue (Asp256^{AUR}) to helix αC and to the activation segment. Mutation of Thr288^{AUR} to Asp generates a protein more active than unphosphorylated wild-type Aurora-A, but much less active than the phosphorylated wild-type kinase. This mutation is thus a poor mimic of phosphorylation at Thr288^{AUR}, and although a Glu may potentially produce a stronger effect by more closely approximating the

substrate-binding site rather than providing a substrate-binding platform. The phosphorylated Thr288^{AUR} has no direct connection with TPX2, the phosphate moiety being 14 Å from the nearest TPX2 atom. Nonetheless, TPX2 induces a 10 Å movement of the phosphoryl moiety required to achieve the active conformation (Figure 4B).

The helical stretch of TPX2 contacts the activation segment at His280^{AUR} and Pro282^{AUR}, which appear to be the pivot points of the conformational change (Figures 4B, 4C). In essence, a rotation about His280^{AUR} swings Pro282^{AUR} and Ser283^{AUR} towards the TPX2 helix, pulling downstream residues of the activation segment to pack more closely with the kinase core (Figure 4B). In this lever-arm-like mechanism, a small change at the pivot points (His280^{AUR}, Pro282^{AUR}) produces a large movement in the arm (P-Thr288^{AUR} Figure 4C). While in the unbound form the activation segment is rather mobile and has probably sufficient conformational freedom to adopt an active conformation in the presence of substrate (basal kinase activity, Figure 1C, lane 1), in the TPX2-bound form, the active conformation is ready for substrate binding and catalysis (increased kinase activity, Figure 1C, lanes 2 and 3). The C-terminal residues (40-43^{TPX}) protruding from a neighbouring TPX2 molecule in the crystals bind to the activation segment, albeit in the opposite main chain direction to a real substrate. Substrate binding alone does not fully activate Aurora-A in solution (Figure 1C), and it is therefore very unlikely that this substrate-like contact is responsible for the active conformation in the crystal structure. Substrate binding, however, might assist in inducing an active conformation, especially in the absence of TPX2.

In addition to enhancing the activity of Aurora-A, TPX2 also protects the kinase from dephosphorylation by phosphatases. The site for PP1 binding on Aurora-A is mapped to the C-terminus, a region that is distant from the TPX2 binding site and that is structurally unaffected upon binding of the regulator. Thus TPX2 is unlikely to protect Aurora-A from dephosphorylation by preventing phosphatase binding. Indeed, even in the presence of TPX2, the phosphatase is capable of removing the phosphates from most side chains with the exception of Thr288^{AUR} (Figure 1E). TPX2 prevents Thr288^{AUR} dephosphorylation by moving the phosphate moiety from a solvent-exposed position to a buried position, which is inaccessible to an incoming phosphatase. The other phosphorylated threonine in the kinase catalytic core, Thr287^{AUR}, is conspicuously exposed to solvent and may serve as a decoy to further mask Thr288^{AUR} (Figure 3B).

Table 1. Summary of crystallographic analysis

		Phospho-Aurora-A + TPX2	Phospho-Aurora-A
Crystals			
Spacegroup		$P2_12_12_1$	$P6_122$
Lattice constants	*a* (Å)	59.63	81.18
	b (Å)	81.72	81.18
	c (Å)	83.05	169.62
Data collection			
X-ray source		SLS XO6SA	ESRF ID14EH1
Resolution range (Å)		40-2.5	70-2.75
(Highest resolution shell)		(2.64-2.5)	(2.9-2.75)
Unique reflections		14609	9235
Completeness(%)		100 (100)	100 (100)
Multiplicity		5.6 (5.7)	10.1 (10.6)
Rmerge (%)		9.9 (24.8)	11.9 (34.6)
I/σ(I)		5.1 (1.1)	3.8 (2.1)
Refinement			
Resolution range (Å)		40-2.5	20-2.75
Number of residues		294	263
Number of waters		144	9
*R*factor (%)		19.4	25.7
*R*free[a] (%)		25.2	29.6
Ramachandran plot			
Most favoured (%)		91.5	82.4
Allowed (%)		8.1	16.3
Generously allowed (%)		0.0	0.9
Forbidden (%)		0.4[b]	0.4[b]

[a] Free *R*factor was computed using 5% of the data assigned randomly (Brunger, 1992).

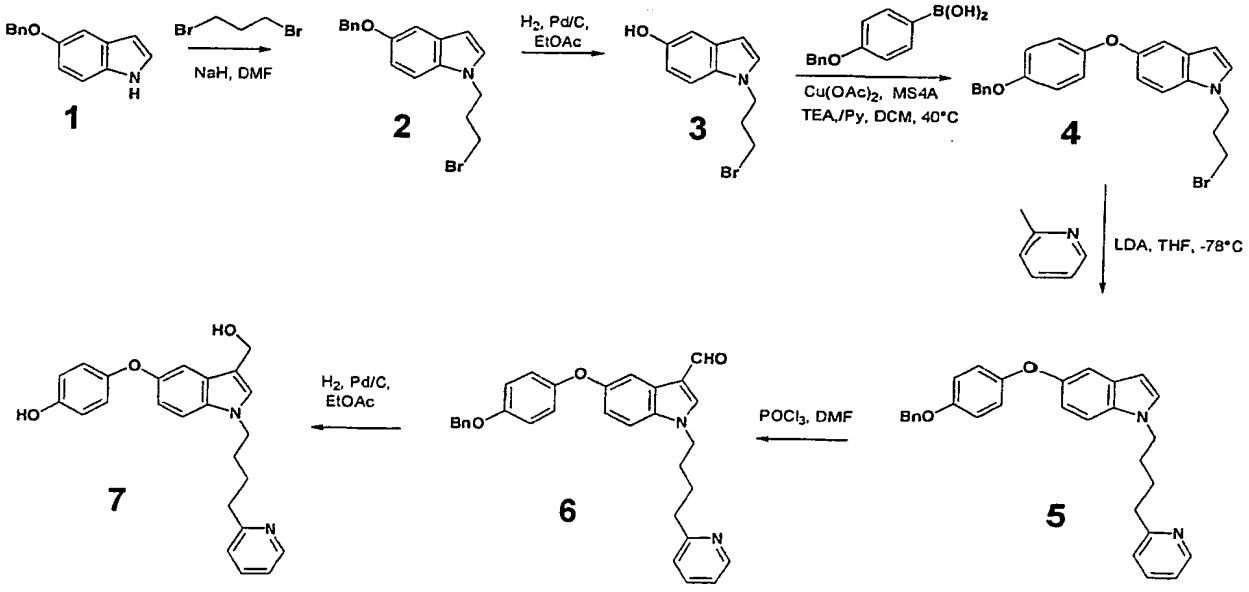
[b] $Ser226^{AUR}$ resides in a loop and the conformation is supported by excellent electron density.

Synthesis of indole compounds of formula (I)

Synthesis of 3-hydroxymethyl-5-O-*p*-hydroxyphenyl-1-N-(4-pyridylbutyl)-indole (7)

As outlined below in Scheme 1, the synthesis started from 5-benzyloxyindole (1). Alkylation of the amino group of (1) with 1,3-dibromopropane was achieved in the presence of NaH in dimethylformamide (DMF) giving the 3'-bromopropyl compound (2). Catalytic hydrogenolysis in the presence of palladium on charcoal (10%) removed the benzyl-protecting group quantitatively. The resulting compound (3) was alkylated at the hydroxy group with a benzyl-protected p-hydroxyphenylboronic acid in the presence of organic base, copper(II)acetate, and molecular sieves (4Å) in dichloromethane (DCM) at 40°C. The resulting 5-O-phenylindole compound (4) was further alkylated to compound (5) by replacing the bromine with 2-methylpyridine in the presence of lithiumdiethylamide (LDA) in tetrahydrofurane (THF) at -78°C. In order to attach more functional groups to the indole ring, (5) was reacted with DMF and phosphorylchloride ($POCl_3$) to give aldehyde (6). Palladium-catalyzed hydrogenolysis in ethylacetate gave target compound (7) in quantitative yield. All compounds were characterized by ^1H- and ^{13}C-NMR as well as mass spectroscopy.

Scheme 1:

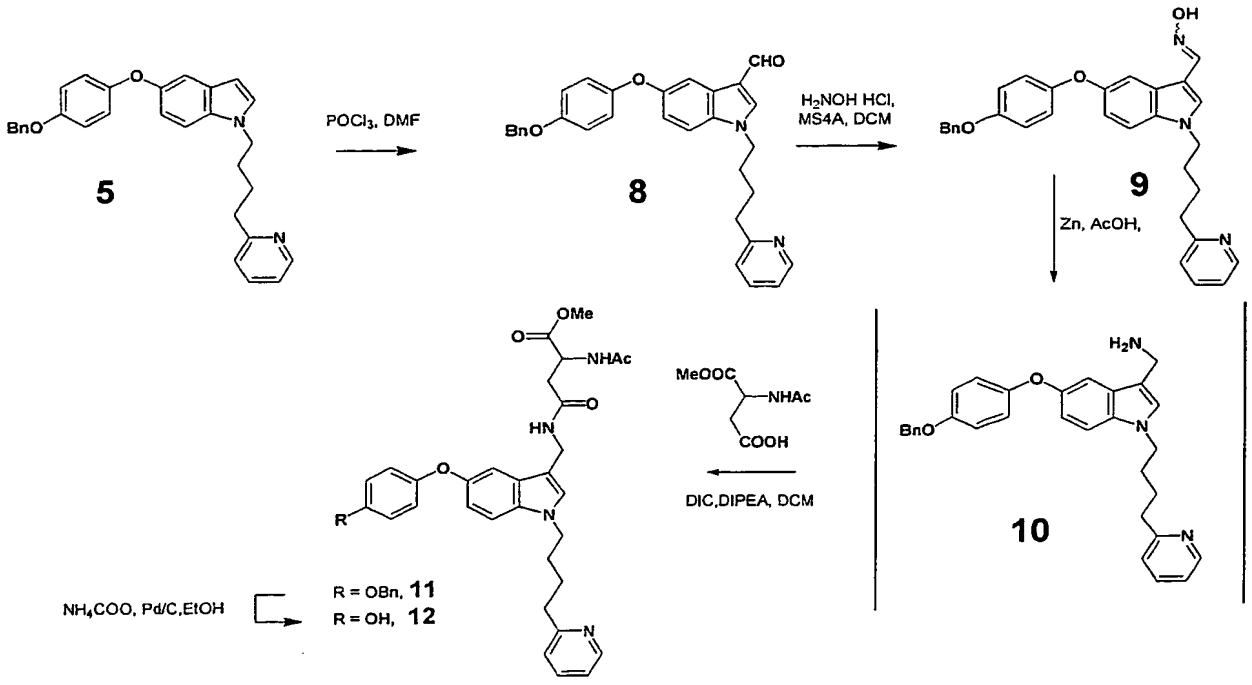


Synthesis of 3-hydroxymethyl indole compound 7.

Synthesis of 3-N-aspartylaminomethyl-5-O-*p*-hydroxyphenyl-1-N-(4-pyridylbutyl)-indole (12)

As outlined below in Scheme 2, the above compound (5) was subjected to Vilsmeyer formylation in DMF/$POCl_3$, which gave aldehyde (8). The latter was reacted with hydroxylamine in DCM in the presence of molecular sieves to give oxime (9). Reduction of the oxime was achieved by zinc in acetic acid. The resulting aminomethyl compound (10) was too labile to be isolated. Instead, the crude mixture was directly reacted with N-acetylaspartyl methylester under standard peptide synthesis conditions (diisopropylcarbodiimide, diisopropylethylamine, DCM). The aspartylaminomethylindole compound (11) was subjected to hydrogenolysis conditions (palladium on charcoal 10%, ethanol) with ammonium formiate as hydrogen donor. The resulting hydroxyphenyl compound (12) was isolated in good yield. All compounds were characterized by ^{1}H- and ^{13}C-NMR as well as mass spectroscopy. Selected data for (12): MS (ESI pos. mode): 581.2 [M+Na$^+$].

Scheme 2:

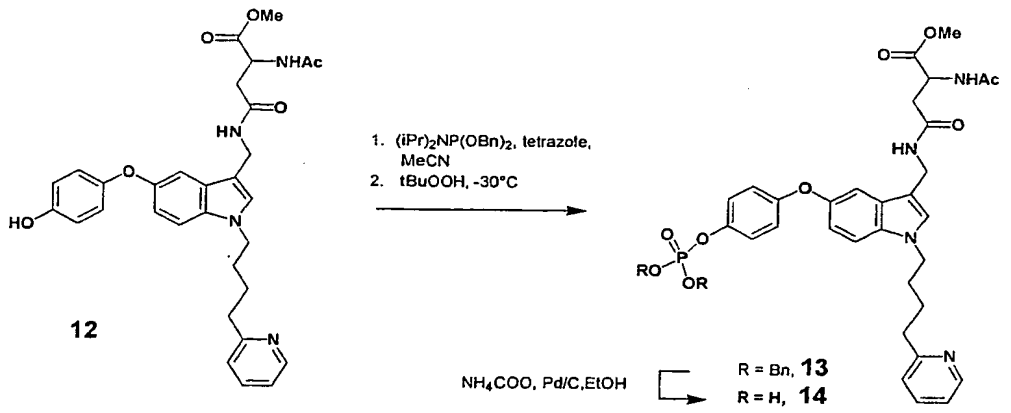


Synthesis of 3-N-aspartylaminomethyl 5-O-p-hydroxyphenyl 1-N-(4-pyridylbutyl) indole compound **12**.

Synthesis of 3-N-aspartylaminomethyl-5-O-*p*-hydroxyphenyl-1-N-(4-pyridylbutyl)-indole phosphate (14)

As outlined below in Scheme 3, the above compound (12) was phosphorylated by standard phosphorus(III) chemistry using dibenzyl-N,N-diisopropylaminophosphoramidite in the presence of dicyanotriazole in acetonitrile under stirring for 12 hours. The mixture was subsequently cooled to -30°C and t-butylperoxyhydroxide was added. The resulting phosphoric acid triester (13) was purified by preparative reversed phase chromatography. Palladium-catalyzed hydrogenolysis quantitatively gave the unprotected 3-N-aspartylaminomethyl-5-O-*p*-hydroxyphenyl-1-N-(4-pyridylbutyl)-indole phosphate (14). All compounds were characterized by ^{1}H-, ^{31}P-, and ^{13}C-NMR as well as mass spectroscopy. Selected data for (14): ^{31}P-NMR: δ = -2.17 ppm. MS (ESI pos. mode): 639.0 [$M+H^{+}$].

Scheme 3:



Phosphorylation of 3-N-aspartylaminomethyl 5-O-p-hydroxyphenyl 1-N-(4-pyridylbutyl) indole compound to the corresponding phosphate ester **14**.

Determination of the inhibition of Aurora A by the compounds of the present invention

The ability of compounds (7) and (12) to inhibit Aurora A was tested in a luminescence ATP detection assay system for kinase applications. Briefly, the compound of the present invention was added to a solution of Aurora A in HEPES buffer (150 nM Aurora +/- TPX2, 25 mM HEPES pH 7.4, 5 mM $MgCl_2$, 0.01% Tween 20; 0.1% BSA; 100 µM Kemptide [Promega Corporation, Madison, WI]); final assay concentrations). The addition of an ATP solution (20 µM final concentration) initiated the reaction. The reaction was carried out for 20 min at room temperature and terminated by the addition of equal volumes of EasyLite Solution (Perkin Elmer Corporation, Boston, MA) followed by monitoring the luminescence in a Wallac Envision plate reader (Perkin Elmer Corporation, Boston, MA). The EasyLite-Kinase® is an

ATP monitoring system based on firely (*Photinus pyralis*) luciferase for the evaluation of kinase activity. The system relies on the consumption of ATP during a kinase reaction and is commercially available.

As can be taken from the following table, both compounds, (7) and (12), inhibited Aurora significantly (but not protein kinase A, PKA).

Compound	Aurora	Aurora + TPX2	PKA
	IC50 (μM)	IC50 (μM)	IC50 (μM)
(7)	79 +/- 4	> 500	> 500
(12)	37 +/- 4	37 +/- 16	> 500